REAL ANALYSIS

REAL ANALYSIS

Modern Techniques and Their Applications

GERALD B. FOLLAND
University of Washington

A WILEY-INTERSCIENCE PUBLICATION
JOHN WILEY & SONS
New York . Chichester . Brisbane . Toronto . Singapore

Copyright © 1984 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work
beyond that permitted by Section 107 or 108 of the
1976 United States Copyright Act without the permission
of the copyright owner is unlawful. Requests for
permission or further information should be addressed to
the Permissions Department, John Wiley & Sons, Inc.

Library of Congress Cataloging in Publication Data:

Folland, G. B.
Real analysis.

"A Wiley-Interscience publication."
Bibliography: p.
Includes index.
1. Mathematical analysis. 2. Functions of real variables. I. Title

QA300.F67 1984 515 84-10435
ISBN 0-471-80958-6

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

For Harold and Helen,

parents and friends

PREFACE

The name "real analysis" is something of an anachronism. Originally applied to the theory of functions of one and several real variables, it has come to encompass several subjects of a more general and abstract nature that underlie much of modern analysis. These general theories and their applications are the subject of this book, which is intended primarily as a text for a graduate-level analysis course. Chapters 1 through 7 are devoted to the core material from measure and integration theory, point set topology, and functional analysis which is a part of most graduate curricula in mathematics, together with a few related but less standard items with which I think all analysts should be acquainted. The last three chapters contain a variety of topics that are meant to introduce some of the other branches of analysis and to illustrate the uses of the preceding material. I believe these topics are all interesting and important, but their selection in preference to others is largely a matter of personal predilection.

The things one needs to know in order to read this book are as follows. (1) First and foremost, the classical theory of functions of real variable: limits and continuity, differentiation and (Riemann) integration, infinite series, uniform convergence, the notion of a metric space. (2) The arithmetic of complex numbers and the basic properties of the complex exponential function $e^{x+iy} = e^x(\cos y + i \sin y)$. (Results from more advanced complex function theory are used only in the proof of the Riesz–Thorin theorem and in a few exercises and remarks.) (3) Some elementary set theory. (4) A bit of linear algebra—actually, not much beyond the definitions of vector spaces, linear mappings, and determinants. All of the necessary material in (1) and (2) can be found, for example, in W. Rudin's *Principles of Mathematical Analysis* (3rd ed., McGraw-Hill, New York, 1976), or T. M. Apostol's *Mathematical Analysis* (2nd ed., Addison-Wesley, Reading, Mass., 1974). A summary of the relevant facts about sets and metric spaces is provided here in the Prologue. The reader should begin this book by examining Sections 1 and 5 of the Prologue to become familiar with my notation and terminology; the rest of the Prologue can then be referred to as needed.

Each chapter concludes with a section entitled "Notes and References." These sections contain miscellaneous remarks, acknowledgments of sources, indications of results not discussed in the text, references for further reading, and historical notes. The latter are quite sketchy, although references to more detailed sources are provided; they are intended mainly to give an idea of how the subject grew out of its classical origins. I found it entertaining and instructive to read some of the original papers and monographs, and I hope to encourage others to do the same.

A sizable portion of this book is devoted to exercises, which are mostly in the form of assertions to be proved and which range from trivial to difficult. Every reader should peruse them, although only the most ambitious will try to work them all out. They serve several purposes: amplification of results and completion of proofs in the text, discussion of examples and counterexamples, applications of theorems, and development of further ideas. Instructors will probably wish to do some of the exercises in class; to maximize flexibility and minimize verbosity, I have followed the principle of "when in doubt, leave it as an exercise." Exercises occur at the end of each section, but they are numbered consecutively within each chapter. In referring to them, "Exercise n" means the nth exercise in the present chapter, while "Exercise $m.n$" means the nth exercise in Chapter m. (The same numbering system is used for sections of chapters).

The topics in the book are arranged so as to allow some flexibility of presentation. For example, Chapters 4 and 5 do not depend on Chapters 1–3 except for a few examples and exercises. On the other hand, if one wishes to proceed quickly to L^p theory, one can skip from Section 3.3 to Sections 5.1 and 5.2 and thence to Chapter 6. The last three chapters are independent of each other, except that the material in Section 8.8 is used in Chapter 9.

The writer of a text on such a well-developed subject as real analysis must necessarily be indebted to his predecessors. I kept a large supply of books on hand while writing this one; they are too numerous to list here, but most of them can be found in the bibliography. I am also happy to acknowledge the influence of two of my teachers: Lynn Loomis, from whose lectures I first learned this subject, and Elias Stein, who has done much to shape my point of view. Finally, I wish to thank my colleagues Edwin Hewitt, Isaac Namioka, Scott Osborne, and Garth Warner for many helpful discussions.

The final draft of this book was written on a VAX 11/750 with the EQN|TROFF text-formatting programs. I am grateful to the people at Bell Laboratories for developing this fine software and to David Ragozin and Douglas Lind for helping me learn how to use it.

GERALD B. FOLLAND

Seattle, Washington
August 1984

CONTENTS

Guide to Notation xiii

Prologue 1

1. The Language of Set Theory, 1
2. Orderings, 5
3. Cardinality, 6
4. More About Well Ordered Sets, 9
5. The Real Number System, 11
6. Metric Spaces, 13
7. Notes and References, 17

1. Measures 18

1. Introduction, 18
2. σ-algebras, 20
3. Measures, 24
4. Outer Measures, 28
5. Borel Measures on **R**, 32
6. Notes and References, 39

2. Integration 42

1. Measurable Functions, 42
2. Integration of Nonnegative Functions, 47
3. Integration of Complex Functions, 51
4. Modes of Convergence, 58
5. Product Measures, 62
6. The Lebesgue Integral on $\mathbf{R}^n$, 68
7. Notes and References, 78

3. Decomposition and Differentiation of Measures **80**

1. Signed Measures, 80
2. The Lebesgue–Radon–Nikodym Theorem, 83
3. Complex Measures, 88
4. Differentiation on $\mathbf{R}^n$, 90
5. Functions of Bounded Variation, 95
6. Notes and References, 105

4. Point Set Topology **107**

1. Topological Spaces, 107
2. Continuous Maps, 113
3. Nets, 119
4. Compact Spaces, 122
5. Locally Compact Spaces, 124
6. Two Compactness Theorems, 129
7. The Stone–Weierstrass Theorem, 132
8. Embeddings in Cubes, 137
9. Notes and References, 141

5. Elements of Functional Analysis **143**

1. Normed Vector Spaces, 143
2. Linear Functionals, 148
3. The Baire Category Theorem and Its Consequences, 153
4. Topological Vector Spaces, 158
5. Hilbert Spaces, 163
6. Notes and References, 171

6. L^p Spaces **173**

1. Basic Theory of L^p Spaces, 173
2. The Dual of L^p, 180
3. Some Useful Inequalities, 185
4. Distribution Functions and Weak L^p, 189
5. Interpolation of L^p Spaces, 192
6. Notes and References, 201

7. Radon Measures **204**

1. Positive Linear Functionals on $C_c(X)$, 204
2. Regularity and Approximation Theorems, 209
3. The Dual of $C_0(X)$, 215
4. Products of Radon Measures, 219
5. Notes and References, 224

8. Topics in Fourier Analysis **226**

1. Preliminaries, 226
2. Convolutions, 230
3. The Fourier Transform, 238
4. Fourier Inversion on the Torus, 249
5. Distributions, 258
6. Sobolev Spaces, 267
7. Applications to Partial Differential Equations, 274
8. Fourier Analysis of Measures, 281
9. Notes and References, 284

9. Topics in Probability Theory **286**

1. Basic Concepts, 286
2. The Law of Large Numbers, 294
3. The Central Limit Theorem, 298
4. Construction of Sample Spaces, 301
5. The Wiener Process, 304
6. Notes and References, 310

10. More Measures and Integrals **312**

1. Topological Groups and Haar Measure, 312
2. Hausdorff Measure, 322
3. Integration on Manifolds, 332
4. Notes and References, 334

Bibliography **335**

Index **341**

GUIDE TO NOTATION

Abstract vector spaces: $L(\mathscr{X}, \mathscr{Y})$ (bounded linear maps), 145. $\mathscr{X}^*$ (dual space), 148.

Analysis on Euclidean space: $f(x^{\pm})$ (one-sided limits), 32. $x'y$ (dot product), 226. ∂^α, x^α, $\alpha!$, $|\alpha|$ (multiindex notation), 227. $\mathbf{T}^n$ (n-torus), 229.

Functions and operations on functions: f^+, f^- (positive and negative parts), 45. sgn, 45. χ_E (characteristic function), 45. f_x, f^y (sections), 63. Γ, 76. supp(f) (support), 125. λ_f (distribution function), 189. f^y (translation), 228. $f * g$ (convolution), 230. f_t (dilation), 233. $\hat{f}, \mathscr{F}f$ (Fourier transform), 240. $\check{f}$ (inverse Fourier transform), 243. $\langle f, \varphi \rangle$, 258. $\tilde{f}$ (reflection), 259.

Integrals: The basic notation is developed in Section 2.2. $\int f(x)\,dx$ (Lebesgue integral), 56, 68. $\int\int f\,d\mu\,d\nu$ (interated integral), 66. $\int g\,dF$ (Stieltjes integral), 99.

Measures: μ_F, 34. m, m^n (Lebesgue measure), 37, 68. $\mu \times \nu$ (product), 62. σ (surface measure on sphere), 74. ν^+, ν^- (positive and negative variations), 82. $|\nu|$ (total variation), 82, 88. $\mu \perp \nu$ (mutual singularity), 82. $\mu \ll \nu$ (absolute continuity), 83. $f\,d\mu$, 84. $d\mu/d\nu$ (Radon-Nikodym derivative), 85. $\mu \hat{\times} \nu$ (Radon product), 221. $\mu * \nu$ (convolution), 281.

Norms and seminorms: $\|f\|_u$ (uniform norm), 115. $\|T\|$ (operator norm), 145. $\|f\|_p$ (L^p norm), 173, 176. $[f]_p$ (weak L^p quasi norm), 191. $\|\mu\|$ (measure norm), 216. $\|f\|_{(s)}$ (Sobolev norm), 268.

Probability theory: $E(X)$ (expectation), 287. $\sigma(X)$ (standard deviation), 287. $\sigma^2(X)$ (variance), 287. P_φ (image measure, distribution), 287. $\nu_\mu^{\sigma^2}$ (normal distribution), 298.

Sets: card(X), 6. $\mathfrak{c}$ (cardinality of the continuum), 8. $F_\sigma, F_{\sigma\delta}, G_\delta, G_{\delta\sigma}$, 21. E_x, E^y (sections), 63.

σ-algebras: $\mathscr{M}(\mathscr{E})$ (σ-algebra generated by $\mathscr{E}$), 21. $\mathscr{B}_X$ (Borel sets), 21. $\otimes_{\alpha \in A} \mathscr{M}_\alpha$, $\mathscr{M} \otimes \mathscr{N}$ (products), 22. $\mathscr{L}, \mathscr{L}^n$ (Lebesgue measurable sets), 37, 68. $\mathscr{B}_X^0$ (Baire sets), 208.

Spaces of functions, measures, and distributions: L^+, 47. L^1, 52. L^1_{loc}, 90. BV, 97. NBV, 98. $C(X, Y)$, 113. $B(X, \mathbf{R})$, 114. $BC(X, \mathbf{R})$, 115. $B(X)$, 115. $C(X)$, 115. $BC(X)$, 115. $C_c(X)$, 125. $C_0(X)$, 126. L^2, 164, 173. l^2, 169, 173. L^p, 173,

176. l^p, 173. weak L^p, 191. $M(X)$, 216. C^k, 226. C^∞, 226. C_c^∞, 226. $\mathscr{S}$, 227. $\mathscr{S}'$, 258. $\mathscr{D}'$, 262. $\mathscr{E}'$, 263. H_s, 268. H_s^{loc}, 271.

Note: For the basic notation used throughout the book for sets, numbers, and metric and topological spaces see Sections 1, 5, and 6 of the Prologue, and Section 4.1. Notation used only in the section in which it is introduced is, for the most part, not listed here.

REAL ANALYSIS

PROLOGUE

The purpose of this introductory chapter is to establish the notation and terminology that will be used throughout the book and to present a few diverse results from set theory and analysis that will be needed later. The style here is deliberately terse, since this chapter is intended as a reference rather than a systematic exposition.

1. THE LANGUAGE OF SET THEORY

It is assumed that the reader is familiar with the basic concepts of set theory; the following brief discussion is meant mainly to fix our terminology.

Logic. We shall avoid the use of special symbols from mathematical logic, preferring to remain reasonably close to standard English. We shall, however, use the abbreviation **iff** for "if and only if."

One point of elementary logic that is often insufficiently appreciated by students is the following. If A and B are mathematical assertions and $-A$, $-B$ are their negations, the statement "A implies B" is logically equivalent to "$-B$ implies $-A$." Thus one may prove that A implies B by assuming $-B$ and deducing $-A$, and we shall frequently do so. This is not the same as *reductio ad absurdum*, which consists of assuming both A and $-B$ and deducing an absurdity.

Number Systems. Our notation for the fundamental number systems is as follows:

$\mathbf{N}$ = the set of positive integers (not including zero)
$\mathbf{Z}$ = the set of integers
$\mathbf{Q}$ = the set of rational numbers
$\mathbf{R}$ = the set of real numbers
$\mathbf{C}$ = the set of complex numbers

Sets. The words "family" and "collection" will be used synonymously with "set," usually to avoid phrases like "set of sets." The empty set is denoted by $\varnothing$, and the family of all subsets of a set X is denoted by $\mathscr{P}(X)$:

$$\mathscr{P}(X) = \{E: E \subset X\},$$

Here and elsewhere, the inclusion sign $\subset$ is interpreted in the weak sense; that is, the assertion "$E \subset X$" includes the possibility that $E = X$.

If $\mathscr{E}$ is a family of sets, we can form the union and intersection of its members:

$$\bigcup_{E\in\mathscr{E}} E = \{x: x \in E \text{ for some } E \in \mathscr{E}\},$$

$$\bigcap_{E\in\mathscr{E}} E = \{x: x \in E \text{ for every } E \in \mathscr{E}\}.$$

Usually it is more convenient to consider indexed families of sets:

$$\mathscr{E} = \{E_\alpha: \alpha \in A\} = \{E_\alpha\}_{\alpha\in A},$$

in which case the union and intersection are denoted by

$$\bigcup_{\alpha\in A} E_\alpha, \qquad \bigcap_{\alpha\in A} E_\alpha.$$

If $E_\alpha \cap E_\beta = \varnothing$ whenever $\alpha \neq \beta$, the sets E_α are called **disjoint**. The terms "disjoint collection of sets" and "collection of disjoint sets" are used interchangeably, as are "disjoint union of sets" and "union of disjoint sets."

When considering families of sets indexed by **N**, our usual notation will be

$$\{E_n\}_{n=1}^{\infty} \quad \text{or} \quad \{E_n\}_1^{\infty},$$

and likewise for unions and intersections. In this situation, the notions of **limit superior** and **limit inferior** are sometimes useful:

$$\limsup E_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} E_n, \qquad \liminf E_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} E_n.$$

The reader may verify that

$$\limsup E_n = \{x: x \in E_n \text{ for infinitely many } n\},$$

$$\liminf E_n = \{x: x \in E_n \text{ for all but finitely many } n\}.$$

If E and F are sets, we denote their **difference** by $E \setminus F$:

$$E \setminus F = \{x: x \in E \text{ and } x \notin F\},$$

and their **symmetric difference** by $E\Delta F$:

$$E\Delta F = (E \setminus F) \cup (F \setminus E).$$

When it is clearly understood that all sets in question are subsets of a fixed set X, we define the **complement** E^c of a set E (in X):

$$E^c = X \setminus E.$$

In this situation we have **deMorgan's laws:**

$$\left(\bigcup_{\alpha\in A} E_\alpha\right)^c = \bigcap_{\alpha\in A} E_\alpha^c, \qquad \left(\bigcap_{\alpha\in A} E_\alpha\right)^c = \bigcup_{\alpha\in A} E_\alpha^c.$$

If X and Y are sets, their **Cartesian product** $X \times Y$ is the set of all ordered pairs (x, y) such that $x \in X$ and $y \in Y$. A **relation** from X to Y is a subset of $X \times Y$. (If $Y = X$, we speak of a relation **on** X.) If R is a relation from X to Y, we shall sometimes write xRy to mean that $(x, y) \in R$. The most important types of relations are the following:

(i) Equivalence relations. An **equivalence relation** on X is a relation R on X such that xRx for all x, xRy iff yRx, and xRz whenever xRy and yRz for some y. The **equivalence class** of $x \in X$ is $\{y \in X: xRy\}$; X is the disjoint union of these equivalence classes.

(ii) Orderings. See Section 2.

(iii) Mappings. A **mapping** $f: X \to Y$ is a relation R from X to Y with the property that for every $x \in X$ there is a unique $y \in Y$ such that xRy, in which case we write $y = f(x)$. Mappings are sometimes called **maps** or **functions**; we shall generally reserve the latter name for the case when Y is **C** or some subset thereof.

If $f: X \to Y$ and $g: Y \to Z$ are mappings, we denote by $g \circ f$ their **composition**:

$$g \circ f: X \to Z, \quad g \circ f(x) = g(f(x)).$$

If $D \subset X$ and $E \subset Y$, we define the **image** of D and the **inverse image** of E under a mapping $f: X \to Y$ by

$$f(D) = \{f(x): x \in D\}, \qquad f^{-1}(E) = \{x: f(x) \in E\}.$$

It is easily verified that the map $f^{-1}: \mathscr{P}(Y) \to \mathscr{P}(X)$ defined by the second

formula commutes with unions, intersections, and complements:

$$f^{-1}\Big(\bigcup_{\alpha\in A} E_\alpha\Big) = \bigcup_{\alpha\in A} f^{-1}(E_\alpha),$$

$$f^{-1}\Big(\bigcap_{\alpha\in A} E_\alpha\Big) = \bigcap_{\alpha\in A} f^{-1}(E_\alpha),$$

$$f^{-1}(E^c) = \big(f^{-1}(E)\big)^c.$$

[The direct image mapping $f\colon \mathscr{P}(X) \to \mathscr{P}(Y)$ commutes with unions, but in general not with intersections and complements.]

If $f\colon X \to Y$ is a mapping, X is called the **domain** of f and $f(X)$ is called the **range** of f. f is said to be **injective** if $f(x_1) = f(x_2)$ only when $x_1 = x_2$, **surjective** if $f(X) = Y$, and **bijective** if it is both injective and surjective. If f is bijective, it has an **inverse** $f^{-1}\colon Y \to X$ such that $f^{-1} \circ f$ and $f \circ f^{-1}$ are the identity mappings on X and Y. If $A \subset X$, we denote by $f|A$ the restriction of f to A:

$$(f|A)\colon A \to Y, \qquad (f|A)(x) = f(x) \quad \text{for } x \in A.$$

A **sequence** in a set X is a mapping from $\mathbf{N}$ into X. (We also use the term **finite sequence** to mean a map from $\{1,\dots,n\}$ into X where $n \in \mathbf{N}$.) If $f\colon \mathbf{N} \to X$ is a sequence and $g\colon \mathbf{N} \to \mathbf{N}$ satisfies $g(n) < g(m)$ whenever $n < m$, the composition $f \circ g$ is called a **subsequence** of f. It is common, and usually convenient, to be careless about distinguishing between sequences and their ranges, which are subsets of X indexed by $\mathbf{N}$. Thus, if $f(n) = x_n$, we speak of the sequence $\{x_n\}_1^\infty$; whether we mean a mapping from $\mathbf{N}$ to X or a subset of X will be clear from the context.

Earlier we defined the Cartesian product of two sets. Similarly one can define the Cartesian product of n sets in terms of ordered n-tuples. However, this definition becomes awkward for infinite families of sets, so the following approach is used instead. If $\{X_\alpha\}_{\alpha\in A}$ is an indexed family of sets, their **Cartesian product** $\prod_{\alpha\in A} X_\alpha$ is the set of all maps $f\colon A \to \bigcup_{\alpha\in A} X_\alpha$ such that $f(\alpha) \in X_\alpha$ for every $\alpha \in A$. (It should be noted, and then promptly forgotten, that when $A = \{1,2\}$, the previous definition of $X_1 \times X_2$ is set-theoretically different from the present definition of $\prod_1^2 X_j$. Indeed, the latter concept depends on mappings, which are defined in terms of the former one.) If $X = \prod_{\alpha\in A} X_\alpha$ and $\alpha \in A$, we define the αth **projection** or **coordinate map** $\pi_\alpha\colon X \to X_\alpha$ by $\pi_\alpha(f) = f(\alpha)$. We also frequently write x and x_α instead of f and $f(\alpha)$ and call x_α the αth **coordinate** of x.

If the sets X_α are all equal to some fixed set Y, then $\prod_{\alpha\in A} X_\alpha$ is simply the set of all mappings from A into Y, and it is denoted by Y^A. If $A = \{1,\dots,n\}$, Y^A is denoted by Y^n and may be identified with the set of ordered n-tuples of elements of Y.

2. ORDERINGS

A **partial ordering** on a nonempty set X is a relation R on X such that (i) if xRy and yRz then xRz; (ii) if xRy and yRx then $x = y$; and (iii) xRx for all x. If R also satisfies (iv) if $x, y \in X$ then either xRy or yRx, then R is called a **linear** (or **total**) **ordering**. For example, if E is any set then $\mathscr{P}(E)$ is partially ordered by inclusion, and $\mathbf{R}$ is linearly ordered by its usual ordering. Taking this last example as a model, we shall usually denote partial orderings by $\le$, and we write $x < y$ to mean that $x \le y$ and $x \ne y$. We observe that a partial ordering on X naturally induces a partial ordering on every nonempty subset of X. Two partially ordered sets X and Y are said to be **order isomorphic** if there is a bijection $f\colon X \to Y$ such that $x_1 \le x_2$ iff $f(x_1) \le f(x_2)$.

If X is partially ordered by $\le$, a **maximal (minimal) element** of X is an element $x \in X$ such that the only $y \in X$ satisfying $x \le y$ ($y \le x$) is x itself. Maximal and minimal elements may or may not exist, and they need not be unique unless the ordering is linear. If $E \subset X$, an **upper (lower) bound** for E is an element $x \in X$ such that $y \le x$ ($x \le y$) for all $y \in E$. An upper bound for E need not be an element of E, and the reader should verify that unless E is linearly ordered, a maximal element of E need not be an upper bound for E.

If X is linearly ordered by $\le$ and every nonempty subset of X has a (necessarily unique) minimal element, X is said to be **well ordered** by $\le$, and, in defiance of the laws of grammar, $\le$ is called a **well ordering** on X. For example, $\mathbf{N}$ is well ordered by its natural ordering.

We now state a fundamental principle of set theory and derive some consequences of it.

(P.1) The Hausdorff Maximal Principle. Every partially ordered set has a maximal linearly ordered subset.

In more detail, this means that if X is partially ordered by $\le$, there is a set $E \subset X$ which is linearly ordered by $\le$, such that no subset of X which properly includes E is linearly ordered by $\le$. Another version of this principle is the following:

(P.2) Zorn's Lemma. If X is a partially ordered set and every linearly ordered subset of X has an upper bound, then X has a maximal element.

Clearly (P.1) implies (P.2): an upper bound for a maximal linearly ordered subset of x is a maximal element of X. It is also not difficult to see that (P.2) implies (P.1). [Apply (P.2) to the collection of linearly ordered subsets of X, which is partially ordered by inclusion.]

(P.3) The Well Ordering Principle. Every nonempty set X can be well ordered.

Proof. Consider the collection $\mathscr{W}$ of well orderings of subsets of X. Such well orderings may be regarded as subsets of $X \times X$, so $\mathscr{W}$ is partially ordered by inclusion. It is easy to verify that the hypotheses of Zorn's lemma are satisfied, so $\mathscr{W}$ has a maximal element. This must be a well ordering of X itself, for if $\leq$ is a well ordering on a proper subset E of X and $x_0 \in X \setminus E$, $\leq$ can be extended to a well ordering on $E \cup \{x_0\}$ by declaring that $x_0 \leq x$ for all $x \in E$. □

(P.4) The Axiom of Choice. If $\{X_\alpha\}_{\alpha \in A}$ is a nonempty collection of nonempty sets, then $\prod_{\alpha \in A} X_\alpha$ is nonempty.

Proof. Let $X = \bigcup_{\alpha \in A} X_\alpha$. Pick a well ordering on X and, for $\alpha \in A$, let $f(\alpha)$ be the minimal element of X_α. Then $f \in \prod_{\alpha \in A} X_\alpha$. □

(P.5) Corollary. If $\{X_\alpha\}_{\alpha \in A}$ is a disjoint collection of nonempty sets, there is a set $Y \subset \bigcup_{\alpha \in A} X_\alpha$ such that $Y \cap X_\alpha$ contains precisely one element for every $\alpha \in A$.

Proof. Take $Y = f(A)$ where $f \in \prod_{\alpha \in A} X_\alpha$. □

We have deduced the axiom of choice from the Hausdorff maximal principle; in fact, the two are known to be logically equivalent.

3. CARDINALITY

If X and Y are nonempty sets, we define the formulas

$$\text{card}(X) \leq \text{card}(Y), \qquad \text{card}(X) = \text{card}(Y), \qquad \text{card}(X) \geq \text{card}(Y)$$

to mean that there exists $f\colon X \to Y$ which is injective, bijective, or surjective, respectively. We also define

$$\text{card}(X) < \text{card}(Y), \qquad \text{card}(X) > \text{card}(Y)$$

to mean that there is an injection but no bijection, or a surjection but no bijection, from X to Y. Observe that we attach no meaning to the expression "card(X)" when it stands alone; there are various ways of doing so, but they are irrelevant for our purposes (except when X is finite—see below). These relationships can be extended to the empty set by declaring that

$$\text{card}(\varnothing) < \text{card}(X) \quad \text{and} \quad \text{card}(X) > \text{card}(\varnothing) \quad \text{for all } X \neq \varnothing.$$

For the remainder of this section we assume implicitly that all sets in question are nonempty in order to avoid special arguments for $\varnothing$. Our first task is to prove that the relationships defined above enjoy the properties which the notation suggests.

(P.6) Proposition. $\text{card}(X) \leq \text{card}(Y)$ iff $\text{card}(Y) \geq \text{card}(X)$.

Proof. If $f\colon X \to Y$ is injective, pick $x_0 \in X$ and define $g\colon Y \to X$ by $g(y) = f^{-1}(y)$ if $y \in f(X)$, $g(y) = x_0$ otherwise. Then g is surjective. Conversely, if $g\colon Y \to X$ is surjective, the sets $g^{-1}(\{x\})$ $(x \in X)$ are nonempty and disjoint, so any $f \in \prod_{x \in X} g^{-1}(\{x\})$ is an injection from X to Y. □

(P.7) Proposition. For any sets X and Y, either $\text{card}(X) \leq \text{card}(Y)$ or $\text{card}(Y) \leq \text{card}(X)$.

Proof. Consider the set $\mathscr{I}$ of all injections from subsets of X into Y. The members of $\mathscr{I}$ can be regarded as subsets of $X \times Y$, so $\mathscr{I}$ is partially ordered by inclusion. It is easily verified that Zorn's lemma applies, so $\mathscr{I}$ has a maximal element f, with (say) domain A and range B. If $x_0 \in X \setminus A$ and $y_0 \in Y \setminus B$, then f can be extended to an injection from $A \cup \{x_0\}$ to $B \cup \{y_0\}$ by setting $f(x_0) = y_0$, contradicting maximality. Hence either $A = X$, in which case $\text{card}(X) \leq \text{card}(Y)$, or $B = Y$, in which case f^{-1} is an injection from Y to X and $\text{card}(Y) \leq \text{card}(X)$. □

(P.8) The Schröder–Bernstein Theorem. If $\text{card}(X) \leq \text{card}(Y)$ and $\text{card}(Y) \leq \text{card}(X)$, then $\text{card}(X) = \text{card}(Y)$.

Proof. Let $f\colon X \to Y$ and $g\colon Y \to X$ be injections. Consider a point $x \in X$: if $x \in \text{range}(g)$ we form $g^{-1}(x) \in Y$; if $g^{-1}(x) \in \text{range}(f)$ we form $f^{-1}(g^{-1}(x)) \in X$; and so forth. Either this process can be continued indefinitely, or it terminates with an element of $X \setminus \text{range}(g)$ (perhaps x itself), or it terminates with an element of $Y \setminus \text{range}(f)$. In these three cases we say that x is in X_∞, X_X, or X_Y: thus X is the disjoint union of X_∞, X_X, and X_Y. In the same way, Y is the disjoint union of three sets Y_∞, Y_X, and Y_Y. Clearly f maps X_∞ onto Y_∞ and X_X onto Y_X, whereas g maps Y_Y onto X_Y. Therefore, if we define $h\colon X \to Y$ by $h(x) = f(x)$ if $x \in X_\infty \cup X_X$ and $h(x) = g^{-1}(x)$ if $x \in X_Y$, then h is bijective. □

(P.9) Proposition. For any set X, $\text{card}(X) < \text{card}(\mathscr{P}(X))$.

Proof. Given $f\colon X \to \mathscr{P}(X)$, let $Y = \{x \in X\colon x \notin f(x)\}$. Then $Y \notin \text{range}(f)$, for if $Y = f(x_0)$ for some $x_0 \in X$, any attempt to answer the question "is $x_0 \in Y$?" quickly leads to an absurdity. Hence no map from X to $\mathscr{P}(X)$ can be surjective. □

A set X is called **countable** if $\text{card}(X) \leq \text{card}(\mathbf{N})$. In particular, all finite sets are countable, and for these it is convenient to interpret "card(X)" as the number of elements in X:

$$\text{card}(X) = n \text{ iff } \text{card}(X) = \text{card}(\{1, \ldots, n\}).$$

(P.10) Proposition. (a) If X is countable and infinite, then $\text{card}(X) = \text{card}(\mathbf{N})$.

(b) If X and Y are countable, so is $X \times Y$.

(c) If A is countable and X_α is countable for every $\alpha \in A$, then $\cup_{\alpha \in A} X_\alpha$ is countable.

Proof. If $E \subset \mathbf{N}$ is infinite, let $f(1)$ be the smallest element of E, and by induction let $f(n)$ be the smallest element of $E \setminus \{f(1), \ldots, f(n-1)\}$. Then f is a bijection from $\mathbf{N}$ to E, from which (a) follows. To prove (b) it suffices to show that $\mathbf{N}^2$ is countable. But we can define a bijection from $\mathbf{N}$ to $\mathbf{N}^2$ by listing those elements $(j,k) \in \mathbf{N}^2$ such that $j + k = n$, for n successively equal to $2, 3, 4, \ldots$, thus:

$$(1,1), (2,1), (1,2), (3,1), (2,2), (1,3), (4,1), (3,2), (2,3), (1,4), \ldots .$$

As for (c), for each $\alpha \in A$ there is a surjective $f_\alpha: \mathbf{N} \to X_\alpha$, and then the map $f: \mathbf{N} \times A \to \cup_{\alpha \in A} X_\alpha$ defined by $f(n, \alpha) = f_\alpha(n)$ is surjective. The result therefore follows from part (b). □

(P.11) Corollary. $\mathbf{Z}$ and $\mathbf{Q}$ are countable.

Proof. $\mathbf{Z}$ is the union of the countable sets $\mathbf{N}$, $\{-n: n \in \mathbf{N}\}$, and $\{0\}$, and the map $f: \mathbf{Z}^2 \to \mathbf{Q}$ defined by $f(m,n) = m/n$ if $n \neq 0$ and $f(m,0) = 0$ is surjective. □

A set X is said to have the **cardinality of the continuum** if $\text{card}(X) = \text{card}(\mathbf{R})$. We shall use the letter $\mathfrak{c}$ as an abbreviation for $\text{card}(\mathbf{R})$:

$$\text{card}(X) = \mathfrak{c} \text{ iff } \text{card}(X) = \text{card}(\mathbf{R}).$$

(P.12) Proposition. $\text{card}(\mathscr{P}(\mathbf{N})) = \mathfrak{c}$.

Proof. If $A \subset \mathbf{N}$, define $f(A) \in \mathbf{R}$ by

$$f(A) = \begin{cases} \sum_{n \in A} 2^{-n} & \text{if } \mathbf{N} \setminus A \text{ is infinite,} \\ 1 + \sum_{n \in A} 2^{-n} & \text{if } \mathbf{N} \setminus A \text{ is finite.} \end{cases}$$

Then f is injective. [If $\mathbf{N} \setminus A$ is infinite, $f(A)$ is the number whose base-2 decimal expansion is $.a_1 a_2 \cdots$, where $a_n = 1$ if $n \in A$ and $a_n = 0$ otherwise.] On the other hand, define $g: \mathscr{P}(\mathbf{Z}) \to \mathbf{R}$ by

$$g(A) = \begin{cases} \log \sum_{n \in A} 2^{-n} & \text{if } A \text{ is bounded below,} \\ 0 & \text{otherwise.} \end{cases}$$

Then g is surjective. Since $\text{card}(\mathscr{P}(\mathbf{N})) = \text{card}(\mathscr{P}(\mathbf{Z}))$, the result follows from the Schröder–Bernstein theorem. □

(P.13) Corollary. If $\text{card}(X) \geq \mathfrak{c}$, then X is uncountable.

Proof. Apply Proposition (P.11). □

The converse of this corollary is the so-called continuum hypothesis; see Section 7.

(P.14) Proposition. (a) If $\text{card}(X) \leq \mathfrak{c}$ and $\text{card}(Y) \leq \mathfrak{c}$, then $\text{card}(X \times Y) \leq \mathfrak{c}$. (b) If $\text{card}(A) \leq \mathfrak{c}$ and $\text{card}(X_\alpha) \leq \mathfrak{c}$ for all $\alpha \in A$, then $\text{card}(\cup_{\alpha \in A} X_\alpha) \leq \mathfrak{c}$.

Proof. For (a) it suffices to take $X = Y = \mathscr{P}(\mathbf{N})$. Define $\varphi, \psi: \mathbf{N} \to \mathbf{N}$ by $\varphi(n) = 2n$ and $\psi(n) = 2n - 1$. It is then easy to check that the map $f: \mathscr{P}(\mathbf{N})^2 \to \mathscr{P}(\mathbf{N})$ defined by $f(A,B) = \varphi(A) \cup \psi(B)$ is a bijection. (b) follows from (a) as in the proof of Proposition (P.10). □

4. MORE ABOUT WELL ORDERED SETS

The material in this section is optional; it is used only in a few exercises and in some notes at the ends of chapters.

Let X be a well ordered set. If $A \subset X$ is nonempty, A has a minimal element, which is its maximal lower bound or **infimum**; we shall denote it by $\inf A$. If A is bounded above, it also has a minimal upper bound or **supremum**, denoted by $\sup A$. If $x \in X$, we define the **initial segment** of x to be

$$I_x = \{y \in X: y < x\}.$$

The elements of I_x are called **predecessors** of x.

The principle of mathematical induction is equivalent to the fact that $\mathbf{N}$ is well ordered. It can be extended to arbitrary well ordered sets as follows:

(P.15) The Principle of Transfinite Induction. Let X be a well ordered set. If A is a subset of X such that $x \in A$ whenever $I_x \subset A$, then $A = X$.

Proof. If $X \neq A$, let $x = \inf(X \setminus A)$. Then $I_x \subset A$ but $x \notin A$. □

(P.16) Proposition. If X is well ordered and $A \subset X$, then $J = \cup_{x \in A} I_x$ is either an initial segment or X itself.

Proof. If $J \neq X$, let $b = \inf(X \setminus J)$. If there existed $y \in J$ with $y > b$, we would have $y \in I_x$ for some $x \in A$ and hence $b \in I_x$, contrary to construction. Hence $J \subset I_x$, and it is obvious that $I_x \subset J$. □

(P.17) Proposition. If X and Y are well ordered, then either X is order isomorphic to Y, or X is order isomorphic to an initial segment in Y, or Y is order isomorphic to an initial segment in X.

Proof. Consider the set $\mathscr{F}$ of order isomorphisms whose domains are initial segments in X or X itself and whose ranges are initial segments in Y or Y itself. $\mathscr{F}$ is nonempty since the unique f: $\{\inf X\} \to \{\inf Y\}$ belongs to $\mathscr{F}$, and $\mathscr{F}$ is partially ordered by inclusion. An application of Zorn's lemma shows that $\mathscr{F}$ has a maximal element f, with (say) domain A and range B. If $A = I_x$ and $B = I_y$, then $A \cup \{x\}$ and $Y \cup \{y\}$ are again initial segments or all of X or Y, and f could be extended by setting $f(x) = y$, contradicting maximality. Hence either $A = X$ or $B = Y$ (or both), and the result follows. □

(P.18) Proposition. There is an uncountable well ordered set Ω such that I_x is countable for each $x \in \Omega$. If Ω' is another set with the same properties, then Ω and Ω' are order isomorphic.

Proof. Uncountable well ordered sets exist by the well ordering principle (P.3); let X be one. Either X has the desired property or there is a minimal element x_0 such that I_{x_0} is uncountable, in which case we can take $\Omega = I_{x_0}$. If Ω' is another such set, Ω' cannot be order isomorphic to an initial segment of Ω or vice versa, because Ω and Ω' are uncountable while their initial segments are countable, so Ω and Ω' are order isomorphic by Proposition (P.17). □

The set Ω in Proposition (P.18), which is essentially unique *qua* well ordered set, is called the **set of countable ordinals.** It has the following remarkable property:

(P.19) Proposition. Every countable subset of Ω has an upper bound.

Proof. If $A \subset \Omega$ is countable, $\bigcup_{x \in A} I_x$ is countable and hence is not all of Ω. By Proposition (P.16), there exists $y \in \Omega$ such that $\bigcup_{x \in A} I_x = I_y$, and y is thus an upper bound for A. □

The set $\mathbf{N}$ of positive integers may be identified with a subset of Ω as follows. Set $f(1) = \inf \Omega$, and by induction set $f(n) = \inf(\Omega \setminus \{f(1), \ldots, f(n-1)\})$. The reader may readily verify that f is an order isomorphism from $\mathbf{N}$ to I_ω, where ω is the minimal element of Ω such that I_ω is infinite.

It is sometimes convenient to add an extra element ω_1 to Ω to form a set $\Omega^* = \Omega \cup \{\omega_1\}$ and to extend the ordering on Ω to Ω^* by declaring that $x < \omega_1$ for all $x \in \Omega$. ω_1 is called the **first uncountable ordinal.** (The usual notation for ω_1 is Ω, since ω_1 is generally taken to be the set of countable ordinals itself.)

5. THE REAL NUMBER SYSTEM

It is frequently useful to adjoin two extra points ∞ $(= +\infty)$ and $-\infty$ to $\mathbf{R}$ to form the **extended real number system** $\overline{\mathbf{R}} = \mathbf{R} \cup \{-\infty, \infty\}$, and to extend the usual ordering on $\mathbf{R}$ to $\overline{\mathbf{R}}$ by declaring that $-\infty < x < \infty$ for all $x \in \mathbf{R}$. The completeness of $\mathbf{R}$ can then be stated as follows: every subset A of $\overline{\mathbf{R}}$ has a least upper bound, or **supremum**, and a greatest lower bound, or **infimum,** which are denoted by $\sup A$ and $\inf A$. If $A = \{a_1, \ldots, a_n\}$, we also write

$$\max(a_1, \ldots, a_n) = \sup A, \qquad \min(a_1, \ldots, a_n) = \inf A.$$

From completeness it follows that every sequence $\{x_n\}$ in $\overline{\mathbf{R}}$ has a **limit superior** and a **limit inferior**:

$$\limsup x_n = \inf_{k \geq 1}(\sup_{n \geq k} x_n), \qquad \liminf x_n = \sup_{k \geq 1}(\inf_{n \geq k} x_n).$$

The sequence $\{x_n\}$ converges iff these two numbers are equal, in which case its limit is their common value. One can also define lim sup and lim inf for functions f: $\mathbf{R} \to \overline{\mathbf{R}}$, for instance:

$$\limsup_{x \to a} f(x) = \inf_{\delta > 0}\big(\sup_{|x-a| < \delta} f(x)\big).$$

The arithmetical operations on $\mathbf{R}$ can be partially extended to $\overline{\mathbf{R}}$:

$$x \pm \infty = \pm\infty \ (x \in \mathbf{R}), \quad \infty + \infty = \infty, \quad -\infty - \infty = -\infty,$$

$$x \cdot (\pm\infty) = \pm\infty \ (x > 0), \quad x \cdot (\pm\infty) = \mp\infty \ (x < 0).$$

We make no attempt to define $\infty - \infty$, but we abide by the convention that, unless otherwise stated,

$$0 \cdot (\pm\infty) = 0.$$

(The expression $0 \cdot \infty$ turns up now and then in measure theory, and for various reasons its proper interpretation is almost always 0.)

We employ the following notation for intervals in $\overline{\mathbf{R}}$: if $-\infty \leq a < b \leq \infty$,

$$(a, b) = \{x: a < x < b\}, \quad [a, b] = \{x: a \leq x \leq b\},$$

$$(a, b] = \{x: a < x \leq b\}, \quad [a, b) = \{x: a \leq x < b\}.$$

We shall occasionally encounter uncountable sums of nonnegative numbers. If X is an arbitrary set and f: $X \to [0, \infty]$, we define $\sum_{x \in X} f(x)$ to be the supremum of its finite partial sums:

$$\sum_{x \in X} f(x) = \sup\Big\{ \sum_{x \in F} f(x): F \subset X, F \text{ finite}\Big\}.$$

(Later we shall recognize this as the integral of f with respect to counting measure on X.)

(P.20) Proposition. Given $f\colon X \to [0,\infty]$, let $A = \{x\colon f(x) > 0\}$. If A is uncountable, then $\Sigma_{x\in X} f(x) = \infty$. If A is countably infinite, then $\Sigma_{x\in X} f(x) = \Sigma_1^\infty f(g(n))$ where $g\colon \mathbf{N} \to A$ is any bijection and the sum on the right is an ordinary infinite series.

Proof. We have $A = \bigcup_1^\infty A_n$ where $A_n = \{x\colon f(x) > 1/n\}$. If A is uncountable then some A_n must be uncountable, so there are arbitrarily large partial sums $\Sigma_{x\in F} f(x)$ where F is a finite subset of A_n. If A is countably infinite, $g\colon \mathbf{N} \to A$ is a bijection, and $B_N = g(\{1,\dots,N\})$, every finite subset of A is contained in some B_N. Hence

$$\sum_{x\in F} f(x) \le \sum_1^N f(g(n)) \le \sum_{x\in X} f(x).$$

Taking the supremum over N, we find

$$\sum_{x\in F} f(x) \le \sum_1^\infty f(g(n)) \le \sum_{x\in X} f(x),$$

and then taking the supremum over F, we obtain the desired result. □

Some terminology concerning (extended) real-valued functions: A relation between numbers which is applied to functions is understood to hold pointwise. Thus, $f \le g$ means that $f(x) \le g(x)$ for every x, and $\max(f,g)$ denotes the function whose value at x is $\max(f(x), g(x))$. If $X \subset \overline{\mathbf{R}}$ and $f\colon X \to \overline{\mathbf{R}}$, f is called **increasing** if $f(x) \le f(y)$ whenever $x \le y$ and **strictly increasing** if $f(x) < f(y)$ whenever $x < y$; similarly for **decreasing**. A function which is either increasing or decreasing is called **monotone**.

For points x in $\mathbf{R}$ or $\mathbf{C}$, $|x|$ denotes the ordinary absolute value or modulus of x. For points $x = (x_1,\dots,x_n)$ in $\mathbf{R}^n$ or $\mathbf{C}^n$, $|x|$ denotes the Euclidean norm:

$$|x| = \left[\sum_1^n |x_j|^2\right]^{1/2}.$$

We recall that a set $U \subset \mathbf{R}$ is **open** if, for every $x \in U$, U includes an interval centered at x.

(P.21) Proposition. Every open set in $\mathbf{R}$ is a countable disjoint union of open intervals.

Proof. If U is open, for each $x \in U$ consider the collection $\mathscr{I}_x$ of all open intervals I such that $x \in I \subset U$, which is nonempty since U is open. It is easy

to check that the union of any family of open intervals containing a point in common is again an open interval, and hence $J_x = \bigcup_{I\in\mathscr{I}_x} I$ is an open interval: it is the largest element of $\mathscr{I}_x$. If $x, y \in U$, then either $J_x = J_y$ or $J_x \cap J_y = \varnothing$, for otherwise $J_x \cup J_y$ would be a larger open interval than J_x in $\mathscr{I}_x$. Thus if $\mathscr{J} = \{J_x\colon x \in U\}$, the (distinct) members of $\mathscr{J}$ are disjoint, and $U = \bigcup_{J\in\mathscr{J}} J$. For each $J \in \mathscr{J}$, pick a rational number $f(J) \in J$. The map $f\colon \mathscr{J} \to \mathbf{Q}$ thus defined is injective, for if $J \ne J'$ then $J \cap J' = \varnothing$; therefore J is countable. □

6. METRIC SPACES

A **metric** on a set X is a function $\rho\colon X \times X \to [0,\infty)$ such that for all $x, y \in X$, (a) $\rho(x,y) = 0$ iff $x = y$, (b) $\rho(x,y) = \rho(y,x)$, and (c) $\rho(x,z) \le \rho(x,y) + \rho(y,z)$. [Intuitively, $\rho(x,y)$ is to be interpreted as the distance from x to y.] A set equipped with a metric is called a **metric space**. Some examples:

(i) The Euclidean distance $\rho(x,y) = |x - y|$ is a metric on $\mathbf{R}^n$.

(ii) $\rho(f,g) = \int_0^1 |f(x) - g(x)|\,dx$ is a metric on the space of continuous functions on $[0,1]$.

(iii) If ρ is a metric on X and $A \subset X$, $\rho|A \times A$ is a metric on A.

(iv) If (X_1,ρ_1) and (X_2,ρ_2) are metric spaces, the **product metric** ρ on $X_1 \times X_2$ is given by

$$\rho((x_1,x_2),(y_1,y_2)) = \max(\rho_1(x_1,y_1), \rho_2(x_2,y_2)).$$

(Other metrics are sometimes used on $X_1 \times X_2$, for instance,

$$\rho_1(x_1,y_1) + \rho_2(x_2,y_2) \quad \text{or} \quad \left[\rho_1(x_1,y_1)^2 + \rho_2(x_2,y_2)^2\right]^{1/2}.$$

These, however, are equivalent to the product metric in the sense which we shall define at the end of the section.)

Let (X,ρ) be a metric space. If $x \in X$ and $r > 0$, the (open) **ball** of radius r about x is

$$B(r,x) = \{y \in X\colon \rho(x,y) < r\}.$$

A set $E \subset X$ is **open** if for every $x \in E$ there exists $r > 0$ such that $B(r,x) \subset E$, and **closed** if its complement E^c is open. For example, every ball $B(r,x)$ is open, for if $y \in B(r,x)$ and $\rho(x,y) = s$ then $B(r-s,y) \subset B(r,x)$. Also, X and $\varnothing$ are both open and closed. Clearly the union of any family of open sets is open, and hence the intersection of any family of closed sets is closed. Also, the intersection (union) of any finite family of open (closed) sets is open (closed). If $E \subset X$, the union of all open sets $U \subset E$ is the largest open set contained in E; it is called the **interior** of E and is denoted by E^o. Likewise, the intersection of all closed sets $F \supset E$ is the smallest closed set

containing E; it is called the **closure** of E and is denoted by $\bar{E}$. E is said to be **dense** (in X) if $\bar{E} = X$, and **nowhere dense** if $\bar{E}$ has empty interior. X is called **separable** if it has a countable dense subset. (For example, $\mathbf{Q}^n$ is a countable dense subset of $\mathbf{R}^n$.) A sequence $\{x_n\}$ in X **converges** to $x \in X$ (symbolically: $x_n \to x$ or $\lim x_n = x$) if $\lim_{n\to\infty} \rho(x_n, x) = 0$.

(P.22) Proposition. If X is a metric space, $E \subset X$, and $x \in X$, the following are equivalent:

(a) $x \in \bar{E}$.

(b) $B(r, x) \cap E \neq \varnothing$ for all $r > 0$.

(c) There is a sequence $\{x_n\}$ in E which converges to x.

Proof. If $B(r, x) \cap E = \varnothing$ then $B(r, x)^c$ is a closed set containing E but not x, so $x \notin \bar{E}$. Conversely, if $x \notin \bar{E}$, since $(\bar{E})^c$ is open there exists $r > 0$ such that $B(r, x) \subset (\bar{E})^c \subset E^c$. Thus (a) is equivalent to (b). If (b) holds, for each $n \in \mathbf{N}$ there exists $x_n \in B(1/n, x) \cap E$, so that $x_n \to x$. On the other hand, if $B(r, x) \cap E = \varnothing$ then $\rho(y, x) \geq r$ for all $y \in E$, so no sequence in E can converge to x. Thus (b) is equivalent to (c). □

If X and Y are metric spaces, with metrics both denoted by ρ, a map $f\colon X \to Y$ is called **continuous at** $x \in X$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\rho(f(y), f(x)) < \varepsilon$ whenever $\rho(x, y) < \delta$—in other words, $f^{-1}(B(\varepsilon, f(x))) \supset B(\delta, x)$. f is **continuous** if it is continuous at each $x \in X$ and **uniformly continuous** if, in addition, the δ in the definition of continuity can be chosen independent of x.

(P.23) Proposition. $f\colon X \to Y$ is continuous iff $f^{-1}(U)$ is open in X for every open $U \subset Y$.

Proof. If the latter condition holds, for every $x \in X$ and $\varepsilon > 0$ $f^{-1}(B(\varepsilon, f(x)))$ is open and contains x, so it contains some ball about x; this means that f is continuous at x. Conversely, suppose f is continuous and U is open in Y. For each $y \in U$ there exists $\varepsilon_y > 0$ such that $B(\varepsilon_y, y) \subset U$, and for each $x \in f^{-1}(\{y\})$ there exists $\delta_x > 0$ such that

$$B(\delta_x, x) \subset f^{-1}\big(B(\varepsilon_y, y)\big) \subset f^{-1}(U).$$

Thus $f^{-1}(U) = \bigcup_{x \in f^{-1}(U)} B(\delta_x, x)$ is open. □

If (X, ρ) is a metric space, a sequence $\{x_n\}$ in X is called **Cauchy** if $\rho(x_n, x_m) \to 0$ as $n, m \to \infty$. A subset E of X is called **complete** if every Cauchy sequence in E converges to a point in E. For example, $\mathbf{R}^n$ (with the Euclidean metric) is complete, while $\mathbf{Q}^n$ is not.

(P.24) Proposition. A closed subset of a complete metric space is complete, and a complete subset of an arbitrary metric space is closed.

Proof. If X is complete, $E \subset X$ is closed, and $\{x_n\}$ is a Cauchy sequence in E, $\{x_n\}$ has a limit in X. By Proposition (P.22), $x \in \bar{E} = E$. If $E \subset X$ is complete and $x \in \bar{E}$, by Proposition (P.22) there is a sequence $\{x_n\}$ in E converging to x. $\{x_n\}$ is Cauchy, so its limit lies in E; thus $E = \bar{E}$. □

In a metric space (X, ρ) we can define the distance from a point to a set and the distance between two sets: if $x \in X$ and $E, F \subset X$,

$$\rho(x, E) = \inf\{\rho(x, y)\colon y \in E\},$$

$$\rho(E, F) = \inf\{\rho(x, y)\colon x \in E, y \in F\} = \inf\{\rho(x, F)\colon x \in E\}.$$

Observe that, by Proposition (P.22), $\rho(x, E) = 0$ iff $x \in \bar{E}$. We also define the **diameter** of $E \subset X$ to be

$$\operatorname{diam} E = \sup\{\rho(x, y)\colon x, y \in E\}.$$

E is called **bounded** if $\operatorname{diam} E < \infty$.

If $E \subset X$ and $\{V_\alpha\}_{\alpha \in A}$ is a family of sets such that $E \subset \bigcup_{\alpha \in A} V_\alpha$, $\{V_\alpha\}_{\alpha \in A}$ is called a **cover** of E, and E is said to be **covered** by the V_α's. E is called **totally bounded** if, for every $\varepsilon > 0$, E can be covered by finitely many balls of radius ε. Every totally bounded set is bounded, for if $x, y \in \bigcup_1^n B(\varepsilon, z_j)$, say $x \in B(\varepsilon, z_1)$ and $y \in B(\varepsilon, z_2)$, then

$$\rho(x, y) \leq \rho(x, z_1) + \rho(z_1, z_2) + \rho(z_2, y)$$

$$\leq 2\varepsilon + \max\{\rho(z_j, z_k)\colon 1 \leq j, k \leq n\}.$$

(The converse is false in general.) If E is totally bounded, so is $\bar{E}$, for it is easily seen that if $E \subset \bigcup_1^n B(\varepsilon, z_j)$ then $\bar{E} \subset \bigcup_1^n B(2\varepsilon, z_j)$.

(P.25) Theorem. If E is a subset of the metric space (X, ρ), the following are equivalent:

(a) E is complete and totally bounded.

(b) (The Bolzano–Weierstrass property). Every sequence in E has a subsequence which converges to a point of E.

(c) (The Heine–Borel property). If $\{V_\alpha\}_{\alpha \in A}$ is a cover of E by open sets, there is a finite set $F \subset A$ such that $\{V_\alpha\}_{\alpha \in F}$ covers E.

Proof. We shall show that (a) and (b) are equivalent, that (a) and (b) together imply (c), and finally that (c) implies (b).

(a) implies (b): Suppose (a) holds and $\{x_n\}$ is a sequence in E. E can be covered by finitely many balls of radius 2^{-1}, and at least one of them must contain x_n for infinitely many n: say, $x_n \in B_1$ for $n \in N_1$. $E \cap B_1$ can be covered by finitely many balls of radius 2^{-2}, and at least one of them must

contain x_n for infinitely many $n \in N_1$: say, $x_n \in B_2$ for $n \in N_2$. Continuing inductively, we obtain a sequence of balls B_j of radius 2^{-j} and a decreasing sequence of subsets N_j of $\mathbf{N}$ such that $x_n \in B_j$ for $n \in N_j$. Pick $n_1 \in N_1$, $n_2 \in N_2, \ldots$, such that $n_1 < n_2 < \cdots$. Then $\{x_{n_j}\}$ is a Cauchy sequence, for $\rho(x_{n_j}, x_{n_k}) < 2^{-j}$ if $k > j$, and since E is complete it has a limit in E.

(b) implies (a): If E is not complete, there is a Cauchy sequence $\{x_n\}$ in E with no limit in E. No subsequence of $\{x_n\}$ can converge in E, for otherwise the whole sequence would converge to the same limit. On the other hand, if E is not totally bounded, let $\varepsilon > 0$ be such that E cannot be covered by finitely many balls of radius ε. Choose $x_n \in E$ inductively as follows: begin with any $x_1 \in E$, and having chosen $x_1, \ldots, x_{n-1}$, pick $x_n \in E \setminus \bigcup_1^{n-1} B(\varepsilon, x_j)$. Then $\rho(x_m, x_n) \geq \varepsilon$ for all m, n, so $\{x_n\}$ has no convergent subsequence.

(a) and (b) imply (c): It suffices to show that if (b) holds and $\{V_\alpha\}_{\alpha \in A}$ is a cover of E by open sets, there exists $\varepsilon > 0$ such that every ball of radius ε which intersects E is contained in some V_α, for E can be covered by finitely many such balls by (a). Suppose to the contrary that for each $n \in \mathbf{N}$ there is a ball B_n of radius $\leq 2^{-n}$ such that $B_n \cap E \neq \varnothing$ and B_n is contained in no V_α. Pick $x_n \in B_n \cap E$; by passing to a subsequence we may assume that $\{x_n\}$ converges to some $x \in E$. We have $x \in V_\alpha$ for some α, and since V_α is open there exists $\varepsilon > 0$ such that $B(\varepsilon, x) \subset V_\alpha$. But if n is large enough so that $\rho(x_n, x) < \varepsilon/3$ and $2^{-n} < \varepsilon/3$, then $B_n \subset B(\varepsilon, x) \subset V_\alpha$, contradicting the assumptions on B_n.

(c) implies (b): If $\{x_n\}$ is a sequence in E with no convergent subsequence, for each $x \in E$ there is a ball B_x centered at x which contains x_n for only finitely many n (otherwise some subsequence would converge to x). Then $\{B_x\}_{x \in E}$ is a cover of E by open sets which has no finite subcover. □

A set E which possesses the properties (a)–(c) of theorem (P.25) is called **compact**. Every compact set is closed [by Proposition (P.24)] and bounded; the converse is false in general but true in $\mathbf{R}^n$:

(P.26) Proposition. Every closed and bounded subset of $\mathbf{R}^n$ is compact.

Proof. Since closed subsets of $\mathbf{R}^n$ are complete, it suffices to show that bounded subsets of $\mathbf{R}^n$ are totally bounded. Since every bounded set is contained in some cube

$$Q = [-R, R]^n = \{x \in \mathbf{R}^n : \max(|x_1|, \ldots, |x_n|) \leq R\},$$

it is enough to show that Q is totally bounded. Given $\varepsilon > 0$, pick an integer $k > \sqrt{n}\, R/\varepsilon$, and express Q as the union of k^n congruent subcubes by dividing the interval $[-R, R]$ into k equal pieces. The side length of these subcubes is $2R/k$ and hence their diameter is $2\sqrt{n}\, R/k < 2\varepsilon$, so they are contained in the balls of radius ε about their centers. □

Two metrics ρ_1 and ρ_2 on a set X are called **equivalent** if

$$C\rho_1 \leq \rho_2 \leq C'\rho_1 \text{ for some } C, C' > 0.$$

It is easily verified that equivalent metrics define the same open, closed, and compact sets, the same convergent and Cauchy sequences, and the same continuous and uniformly continuous mappings. Consequently, most results concerning metric spaces depend not on the particular metric chosen but only on its equivalence class.

7. NOTES AND REFERENCES

Sections 1–4. More detailed accounts of this material, including the deduction of the Hausdorff maximal principle from the axiom of choice, can be found in Halmos [42], Hewitt and Stromberg [54], and Kelley [58].

The axiom of choice (or one of the propositions equivalent to it) is generally taken as one of the basic postulates in the axiomatic formulations of set theory. Some mathematicians of the intuitionist or constructivist persuasion reject it on the grounds that one has not proved the existence of a mathematical object until one has shown how to construct it in some reasonably explicit fashion, whereas the whole point of the axiom of choice is to provide existence theorems when constructive methods fail (or are too cumbersome for comfort). People who are seriously bothered by such objections belong to a minority which does not include the present writer; in this book the axiom of choice is used sparingly but freely.

The **continuum hypothesis** is the assertion that if $\operatorname{card}(X) < \mathfrak{c}$ then X is countable. [Since it follows easily from the construction of Ω, the set of countable ordinals, that $\operatorname{card}(\Omega) \leq \operatorname{card}(X)$ for any uncountable X, an equivalent assertion is that $\operatorname{card}(\Omega) = \mathfrak{c}$.] It is known, thanks to Gödel and Cohen, that the continuum hypothesis and its negation are both consistent with the standard axioms of set theory including the axiom of choice (assuming those axioms are consistent): see Cohen [18]. Some mathematicians are willing to accept the continuum hypothesis as true, seemingly as a matter of convenience, but Gödel [38] and Cohen [18] have both expressed suspicions that it should be false, and as of this writing no one has found any really compelling evidence on one side or the other. My own feeling, subject to revision in the event of a major breakthrough in set theory, is that if the answer to one's question turns out to depend on the continuum hypothesis, one should give up and ask a different question.

Section 6. A fuller discussion of metric spaces can be found in Loomis and Sternberg [66], Chapter 4.

1

MEASURES

In this chapter we set forth the basic concepts of measure theory, develop a general procedure for constructing nontrivial examples of measures, and apply this procedure to construct measures on the real line.

1. INTRODUCTION

One of the most venerable problems in geometry is to determine the area or volume of a region in the plane or in 3-space. The techniques of integral calculus provide a satisfactory solution to this problem for regions which are bounded by "nice" curves or surfaces but are inadequate to handle more complicated sets, even in dimension one. Ideally, for each $n \in \mathbf{N}$ we would like to have a function μ which assigns to each $E \subset \mathbf{R}^n$ a number $\mu(E) \in [0, \infty]$, the n-dimensional measure of E, such that $\mu(E)$ is given by the usual integral formulas when the latter apply. Such a function μ should surely possess the following properties:

(i) If $E_1, E_2, \ldots$ is a finite or infinite sequence of disjoint sets, then

$$\mu(E_1 \cup E_2 \cup \cdots) = \mu(E_1) + \mu(E_2) + \cdots .$$

(ii) If E is congruent to F (that is, if E can be transformed into F by translations, rotations, and reflections), then $\mu(E) = \mu(F)$.

(iii) $\mu(Q) = 1$, where Q is the unit cube:

$$Q = \{x \in \mathbf{R}^n : 0 \le x_j < 1 \text{ for } j = 1, \ldots, n\}.$$

Unfortunately, these conditions are mutually inconsistent. Let us see why this is true for $n = 1$. (The argument can easily be adapted to higher dimensions.) To begin with, we define an equivalence relation on $[0,1)$ by declaring

that $x \sim y$ iff $x - y$ is rational. Let N be a subset of $[0,1)$ which contains precisely one member of each equivalence class. (Finding such an N requires invoking the axiom of choice.) Next, let $R = \mathbf{Q} \cap [0,1)$, and for each $r \in R$ let

$$N_r = \{x + r : x \in N \cap [0, 1 - r)\} \cup \{x + r - 1 : x \in N \cap [1 - r, 1)\}.$$

That is, to obtain N_r, move N r units to the right and then move the part which sticks out beyond $[0,1)$ one unit to the left. Then every $x \in [0,1)$ belongs to precisely one N_r. Indeed, if y is the element of N which belongs to the equivalence class of x, then $x \in N_r$ where $r = x - y$ if $x \ge y$ or $r = x - y + 1$ if $x < y$; on the other hand, if $x \in N_r \cap N_s$, then $x - r$ (or $x - r + 1$) and $x - s$ (or $x - s + 1$) would be distinct elements of N belonging to the same equivalence class, which is impossible.

Suppose now that μ: $\mathscr{P}(\mathbf{R}) \to [0, \infty]$ satisfies (i), (ii), and (iii). By (i) and (ii),

$$\mu(N) = \mu(N \cap [0, 1 - r)) + \mu(N \cap [1 - r, 1)) = \mu(N_r)$$

for any $r \in R$, and since R is countable,

$$\mu([0,1)) = \sum_{r \in R} \mu(N_r).$$

But $\mu([0,1)) = 1$ by (iii), and since $\mu(N_r) = \mu(N)$, the sum on the right is either 0 (if $\mu(N) = 0$) or ∞ (if $\mu(N) > 0$). Hence no such μ can exist.

Faced with this discouraging situation, one might consider weakening (i) so that additivity is required to hold only for finite sequences. This is not a very good idea, as we shall see: the additivity for countable sequences is what makes all the limit and continuity results of the theory work smoothly. Moreover, in dimensions $n \ge 3$, even this weak form of (i) is inconsistent with (ii) and (iii). Indeed, in 1924 Banach and Tarski proved the following amazing result:

Let U and V be bounded open sets in $\mathbf{R}^n$, $n \ge 3$. There exist $k \in \mathbf{N}$ and subsets $E_1, \ldots, E_k$, $F_1, \ldots, F_k$ of $\mathbf{R}^n$ such that: (a) the E_j's are disjoint and their union is U, (b) the F_j's are disjoint and their union is V, and (c) E_j is congruent to F_j for $j = 1, \ldots, k$.

Thus one can cut up a ball the size of a pea into a finite number of pieces and rearrange the pieces to form a ball the size of the earth! Needless to say, the sets E_j and F_j are *very* bizarre. They cannot be visualized accurately, and their construction depends on the axiom of choice. But their existence clearly precludes the construction of any μ: $\mathscr{P}(\mathbf{R}) \to [0, \infty]$ satisfying (ii) and the weakened form of (i) for which bounded open sets have positive, finite measure.

The moral of these examples is that $\mathbf{R}^n$ contains subsets which are so strangely put together that it is impossible to define a geometrically reasonable

notion of measure for them, and the remedy for the situation is to discard the requirement that μ should be defined on all subsets of $\mathbf{R}^n$. Rather, we shall content ourselves with constructing μ on a class of subsets of $\mathbf{R}^n$ which includes all the sets one is likely to meet in practice unless one is deliberately searching for pathological examples. This construction will be carried out for $n = 1$ in Section 5 (Chapter 1) and for $n > 1$ in Section 2.6.

It is worthwhile, and not much extra work, to develop the theory in much greater generality. The conditions (ii) and (iii) are directly related to Euclidean geometry, but set functions satisfying (i), called *measures*, arise also in a great many other situations. For example, in a physics problem involving mass distributions, $\mu(E)$ could be the total mass in the region E. For another example, in probability theory one considers a set X which represents the possible outcomes of an experiment, and for $E \subset X$, $\mu(E)$ is the probability that the outcome lies in E. We therefore begin by studying the theory of measures on abstract sets.

2. σ-ALGEBRAS

In this section we discuss the families of sets which are the domains of measures.

Let X be a nonempty set. An **algebra** of sets on X is a nonempty collection $\mathscr{A} \subset \mathscr{P}(X)$ which is closed under finite unions and complements: that is, if $E_1, \ldots, E_n \in \mathscr{A}$ then $\bigcup_1^n E_j \in \mathscr{A}$, and if $E \in \mathscr{A}$ then $E^c \in \mathscr{A}$. A **σ-algebra** is an algebra which is closed under countable unions. (Some authors use the terms **field** and **σ-field** instead of algebra and σ-algebra.) We observe that since $\bigcap_j E_j = (\bigcup_j E_j^c)^c$, algebras (resp. σ-algebras) are also closed under finite (resp. countable) intersections. Moreover, if $\mathscr{A}$ is an algebra then $\varnothing \in \mathscr{A}$ and $X \in \mathscr{A}$, for given any $E \in \mathscr{A}$, $\varnothing = E \cap E^c$ and $X = E \cup E^c$.

It is worth noting that an algebra $\mathscr{A}$ is a σ-algebra provided that it is closed under countable *disjoint* unions. Indeed, suppose $\{E_j\}_1^\infty \subset \mathscr{A}$. Set

$$F_k = E_k \Big\backslash \left[\bigcup_1^{k-1} E_j\right] = E_k \cap \left(\bigcup_1^{k-1} E_j\right)^c.$$

Then the F_k's belong to $\mathscr{A}$ and are disjoint, and $\bigcup_1^\infty E_j = \bigcup_1^\infty F_k$. This device of replacing a sequence of sets by a disjoint sequence is worth remembering; it will be used a number of times below.

Some examples: If X is any set, $\mathscr{P}(X)$ and $\{\varnothing, X\}$ are σ-algebras. If X is uncountable, then

$$\mathscr{A} = \{E \subset X\colon E \text{ is countable or } E^c \text{ is countable}\}$$

is a σ-algebra, called the **σ-algebra of countable or co-countable sets**. (The point here is that if $\{E_j\}_1^\infty \subset \mathscr{A}$, then $\bigcup_1^\infty E_j$ is countable if all E_j are countable and is co-countable otherwise.)

It is trivial to verify that the intersection of any family of σ-algebras on X is again a σ-algebra. It follows that if $\mathscr{E}$ is any subset of $\mathscr{P}(X)$, there is a unique smallest σ-algebra $\mathscr{M}(\mathscr{E})$ containing $\mathscr{E}$, namely, the intersection of all σ-algebras containing $\mathscr{E}$. [There is always at least one such, namely $\mathscr{P}(X)$.] $\mathscr{M}(\mathscr{E})$ is called the σ-algebra **generated** by $\mathscr{E}$. The following observation is often useful:

(1.1) Lemma. If $\mathscr{E} \subset \mathscr{M}(\mathscr{F})$, then $\mathscr{M}(\mathscr{E}) \subset \mathscr{M}(\mathscr{F})$.

Proof. $\mathscr{M}(\mathscr{F})$ is a σ-algebra containing $\mathscr{E}$; it therefore contains $\mathscr{M}(\mathscr{E})$. □

If X is any metric space, or more generally any topological space (see Chapter 4), the σ-algebra generated by the family of open sets in X (or, equivalently, by the family of closed sets in X) is called the **Borel σ-algebra** on X and will be denoted by $\mathscr{B}_X$. Its members are called **Borel sets**. $\mathscr{B}_X$ thus includes open sets, closed sets, countable intersections of open sets, countable unions of closed sets, and so forth. It is convenient to have a terminology for the levels in this hierarchy: a countable intersection of open sets is called a G_δ set, a countable union of closed sets is called an F_σ set, a countable union of G_δ sets is called a $G_{\delta\sigma}$ set, a countable intersection of F_σ sets is called an $F_{\sigma\delta}$ set, etc. (δ and σ stand for the German *Durchschnitt* and *Summe*, that is, intersection and union.)

The Borel σ-algebra on $\mathbf{R}$ will play a fundamental role in what follows. For future reference we note that it can be generated in a number of different ways:

(1.2) Proposition. $\mathscr{B}_\mathbf{R}$ is generated by each of the following:

(a) the open intervals: $\mathscr{E}_1 = \{(a, b)\colon a < b\}$

(b) the closed intervals: $\mathscr{E}_2 = \{[a, b]\colon a < b\}$

(c) the half-open intervals: $\mathscr{E}_3 = \{(a, b]\colon a < b\}$ or $\mathscr{E}_4 = \{[a, b)\colon a < b\}$

(d) the open rays: $\mathscr{E}_5 = \{(a, \infty)\colon a \in \mathbf{R}\}$ or $\mathscr{E}_6 = \{(-\infty, a)\colon a \in \mathbf{R}\}$

(e) the closed rays: $\mathscr{E}_7 = \{[a, \infty)\colon a \in \mathbf{R}\}$ or $\mathscr{E}_8 = \{(-\infty, a]\colon a \in \mathbf{R}\}$.

Proof. The elements of $\mathscr{E}_j$ for $j \neq 3, 4$ are open or closed, and the elements of $\mathscr{E}_3$ and $\mathscr{E}_4$ are G_δ sets (e.g., $(a, b] = \bigcap_1^\infty (a, b + n^{-1})$). All of these are Borel sets, so by Lemma (1.1), $\mathscr{M}(\mathscr{E}_j) \subset \mathscr{B}_\mathbf{R}$ for all j. On the other hand, every open set in $\mathbf{R}$ is a countable union of open intervals, so by Lemma (1.1) again, $\mathscr{B}_\mathbf{R} \subset \mathscr{M}(\mathscr{E}_1)$. That $\mathscr{B}_\mathbf{R} \subset \mathscr{M}(\mathscr{E}_j)$ for $j \geq 2$ can now be established by showing that all open intervals lie in $\mathscr{M}(\mathscr{E}_j)$ and applying Lemma (1.1). For example, $(a, b) = \bigcup_1^\infty [a + n^{-1}, b - n^{-1}] \in \mathscr{M}(\mathscr{E}_2)$. Verification of the other cases is left to the reader (Exercise 2). □

Let $\{X_\alpha\}_{\alpha\in A}$ be an indexed collection of nonempty sets, $X = \prod_{\alpha\in A} X_\alpha$, and $\pi_\alpha\colon X \to X_\alpha$ the coordinate maps. If $\mathscr{M}_\alpha$ is a σ-algebra on X_α for each α, the **product σ-algebra** on X is the σ-algebra generated by

$$\{\pi_\alpha^{-1}(E_\alpha)\colon E_\alpha \in \mathscr{M}_\alpha, \alpha \in A\}.$$

We denote this σ-algebra by $\otimes_{\alpha\in A}\mathscr{M}_\alpha$. (If $A = \{1,\dots,n\}$ we also write $\otimes_1^n \mathscr{M}_j$ or $\mathscr{M}_1 \otimes \cdots \otimes \mathscr{M}_n$.) The significance of this definition will become clearer in Section 2.1; for the moment we give an alternative, and perhaps more intuitive, characterization of product σ-algebras in the case of countably many factors.

(1.3) Proposition. If A is countable, then $\otimes_{\alpha\in A}\mathscr{M}_\alpha$ is the σ-algebra generated by $\{\prod_{\alpha\in A} E_\alpha\colon E_\alpha \in \mathscr{M}_\alpha\}$.

Proof. If $E_\alpha \in \mathscr{M}_\alpha$, then $\pi_\alpha^{-1}(E_\alpha) = \prod_{\beta\in A} E_\beta$ where $E_\beta = X_\beta$ for $\beta \neq \alpha$; on the other hand, $\prod_{\alpha\in A} E_\alpha = \bigcap_{\alpha\in A}\pi_\alpha^{-1}(E_\alpha)$. The result therefore follows from Lemma (1.1). □

(1.4) Proposition. Suppose $\mathscr{M}_\alpha$ is generated by $\mathscr{E}_\alpha$, $\alpha \in A$. Then $\otimes_{\alpha\in A}\mathscr{M}_\alpha$ is generated by $\mathscr{F}_1 = \{\pi_\alpha^{-1}(E_\alpha);\ E_\alpha \in \mathscr{E}_\alpha, \alpha \in A\}$. If A is countable and $X_\alpha \in \mathscr{E}_\alpha$ for all α, $\otimes_{\alpha\in A}\mathscr{M}_\alpha$ is also generated by $\mathscr{F}_2 = \{\prod_{\alpha\in A} E_\alpha\colon E_\alpha \in \mathscr{E}_\alpha\}$.

Proof. Obviously $\mathscr{M}(\mathscr{F}_1) \subset \otimes_{\alpha\in A}\mathscr{M}_\alpha$. On the other hand, for each α, $\{E \subset X_\alpha\colon \pi_\alpha^{-1}(E) \in \mathscr{M}(\mathscr{F}_1)\}$ is easily seen to be a σ-algebra on X_α which contains $\mathscr{E}_\alpha$ and hence $\mathscr{M}_\alpha$. In other words, $\pi_\alpha^{-1}(E) \in \mathscr{M}(\mathscr{F}_1)$ for all $E \in \mathscr{M}_\alpha$, $\alpha \in A$, and hence $\otimes_{\alpha\in A}\mathscr{M}_\alpha \subset \mathscr{M}(\mathscr{F}_1)$. The second assertion follows from the first as in the proof of Proposition (1.3). □

(1.5) Proposition. Let $X_1,\dots,X_n$ be metric spaces and let $X = \prod_1^n X_j$, equipped with the product metric. Then $\otimes_1^n \mathscr{B}_{X_j} \subset \mathscr{B}_X$. If the X_j's are separable, then $\otimes_1^n \mathscr{B}_{X_j} = \mathscr{B}_X$.

Proof. By Proposition (1.4), $\otimes_1^n \mathscr{B}_{X_j}$ is generated by the sets $\pi_j^{-1}(U_j)$, $1 \le j \le n$, where U_j is open in X_j. Since these sets are open in X, Lemma (1.1) implies that $\otimes_1^n \mathscr{B}_{X_j} \subset \mathscr{B}_X$. Suppose now that $\{x_j^k\}_{k=1}^\infty$ is a countable dense set in X_j, $1 \le j \le n$, and let $\mathscr{E}_j$ be the collection of balls with rational radius and center at some x_j^k. Then every open set in X_j is a union of members of $\mathscr{E}_j$—in fact, a countable union since $\mathscr{E}_j$ itself is countable. Moreover, the set of points in X whose coordinates are among the x_j^k's is a countable dense subset of X, and the balls of radius r in X are merely products of balls of radius r in the X_j's. It follows that $\mathscr{B}_{X_j}$ is generated by $\mathscr{E}_j$ and $\mathscr{B}_X$ is generated by $\{E_1 \times \cdots \times E_n\colon E_j \in \mathscr{E}_j\}$. Therefore $\mathscr{B}_X = \otimes_1^n \mathscr{B}_{X_j}$ by Proposition (1.4). □

(1.6) Corollary. $\mathscr{B}_{\mathbf{R}^n} = \otimes_1^n \mathscr{B}_{\mathbf{R}}$.

We conclude this section with a technical result which will be needed later. We define an **elementary family** to be a collection $\mathscr{E}$ of subsets of X such that

(i) $\varnothing \in \mathscr{E}$; (ii) if $E, F \in \mathscr{E}$ then $E \cap F \in \mathscr{E}$; and (iii) if $E \in \mathscr{E}$ then E^c is a finite disjoint union of members of $\mathscr{E}$.

(1.7) Proposition. If $\mathscr{E}$ is an elementary family, the collection $\mathscr{A}$ of finite disjoint unions of members of $\mathscr{E}$ is an algebra.

Proof. For simplicity of notation we shall assume that if $E \in \mathscr{E}$ then E^c is a disjoint union of two members of $\mathscr{E}$ (which includes the case $E^c \in \mathscr{E}$, since $E^c = E^c \cup \varnothing$); the proof in the general case is essentially the same. If $A, B \in \mathscr{E}$ and $B^c = C_1 \cup C_2$ ($C_1, C_2 \in \mathscr{E}$, disjoint), then $A \setminus B = (A \cap C_1) \cup (A \cap C_2)$ and $A \cup B = (A \setminus B) \cup B$, where these unions are disjoint, so $A \cup B \in \mathscr{A}$. It now follows by induction that if $A_1,\dots,A_n \in \mathscr{E}$ then $\bigcup_1^n A_j \in \mathscr{A}$: indeed, by inductive hypothesis we may assume that $A_1,\dots,A_{n-1}$ are disjoint, and then $\bigcup_1^n A_j = (\bigcup_1^{n-1}(A_j \setminus A_n)) \cup A_n$, which is a disjoint union. $\mathscr{A}$ is therefore closed under finite unions. To show that $\mathscr{A}$ is closed under complements, suppose $A_1,\dots,A_n \in \mathscr{E}$ and $A_j^c = B_j^1 \cup B_j^2$ ($B_j^1, B_j^2 \in \mathscr{E}$, disjoint). Then

$$\left(\bigcup_1^n A_j\right)^c = \bigcap_1^n \left(B_j^1 \cup B_j^2\right) = \bigcup\left\{B_1^{k_1} \cap \cdots \cap B_n^{k_n}\colon k_1,\dots,k_n = 1,2\right\} \in \mathscr{A}.$$

□

EXERCISES

1. A family of sets $\mathscr{R} \subset \mathscr{P}(X)$ is called a **ring** if it is closed under finite unions and differences (i.e., if $E_1,\dots,E_n \in \mathscr{R}$ then $\bigcup_1^n E_j \in \mathscr{R}$ and if $E, F \in \mathscr{R}$ then $E \setminus F \in \mathscr{R}$). A ring which is closed under countable unions is called a **σ-ring.**
   (a) Rings (resp. σ-rings) are closed under finite (resp. countable) intersections.
   (b) If $\mathscr{R}$ is a ring (resp. σ-ring), then $\mathscr{R}$ is an algebra (resp. σ-algebra) iff $X \in \mathscr{R}$.
   (c) If $\mathscr{R}$ is a σ-ring, then $\{E \subset X\colon E \in \mathscr{R} \text{ or } E^c \in \mathscr{R}\}$ is a σ-algebra.
   (d) If $\mathscr{R}$ is a σ-ring, then $\{E \subset X\colon E \cap F \in \mathscr{R} \text{ for all } F \in \mathscr{R}\}$ is a σ-algebra.
2. Complete the proof of Proposition 1.2.
3. Let $\mathscr{M}$ be an infinite σ-algebra.
   (a) $\mathscr{M}$ contains an infinite sequence of disjoint nonempty sets.
   (b) $\operatorname{card}(\mathscr{M}) \ge \mathfrak{c}$.
4. An algebra $\mathscr{A}$ is a σ-algebra iff $\mathscr{A}$ is closed under countable increasing unions (i.e., if $\{E_j\}_1^\infty \subset \mathscr{A}$ and $E_1 \subset E_2 \subset \cdots$, then $\bigcup_1^\infty E_j \in \mathscr{A}$).

5. If $\mathscr{M}$ is the σ-algebra generated by $\mathscr{E}$, then $\mathscr{M}$ is the union of the σ-algebras generated by $\mathscr{F}$ as $\mathscr{F}$ ranges over all countable subsets of $\mathscr{E}$. (Hint: show that the latter object is a σ-algebra.)

3. MEASURES

Let X be a set equipped with a σ-algebra $\mathscr{M}$. A **measure** on $\mathscr{M}$ [or on $(X, \mathscr{M})$, or simply on X if $\mathscr{M}$ is understood] is a function $\mu\colon \mathscr{M} \to [0, \infty]$ such that

(i) $\mu(\varnothing) = 0$,

(ii) if $\{E_j\}_1^\infty$ is a sequence of disjoint sets in $\mathscr{M}$, then $\mu(\bigcup_1^\infty E_j) = \sum_1^\infty \mu(E_j)$.

Property (ii) is called **countable additivity**. It implies **finite additivity**:

(ii′) if $E_1, \dots, E_n$ are disjoint sets in $\mathscr{M}$, then $\mu(\bigcup_1^n E_j) = \sum_1^n \mu(E_j)$,

because one can take $E_j = \varnothing$ for $j > n$. A function μ which satisfies (i) and (ii′) but not necessarily (ii) is called a **finitely additive measure**.

Here follows a list of the standard terminology concerning measures. If X is a set and $\mathscr{M} \subset \mathscr{P}(X)$ is a σ-algebra, $(X, \mathscr{M})$ is called a **measurable space** and the sets in $\mathscr{M}$ are called **measurable sets**. If μ is a measure on $(X, \mathscr{M})$, then $(X, \mathscr{M}, \mu)$ is called a **measure space**. [The notation $(X, \mathscr{M}, \mu)$ is actually redundant: if one knows μ, one knows $\mathscr{M}$, which is the domain of μ; and if one knows $\mathscr{M}$, one knows X, which is the maximal element of $\mathscr{M}$. Nonetheless, for reasons of clarity one often specifies X and $\mathscr{M}$ explicitly.] If $\mu(X) < \infty$ [which implies that $\mu(E) < \infty$ for all $E \in \mathscr{M}$], μ is called **finite**. If $X = \bigcup_1^\infty E_j$ where $E_j \in \mathscr{M}$ and $\mu(E_j) < \infty$ for all j, μ is called **σ-finite**. More generally, if $E = \bigcup_1^\infty E_j$ where $E_j \in \mathscr{M}$ and $\mu(E_j) < \infty$ for all j, the set E is said to be **σ-finite** (for μ). (It would be more correct, but also more cumbersome, to say that E is of σ-finite measure.) Finally, if for each $E \in \mathscr{M}$ with $\mu(E) = \infty$ there exists $F \in \mathscr{M}$ with $F \subset E$ and $0 < \mu(F) < \infty$, μ is called **semifinite**.

Every σ-finite measure is semifinite, but not conversely (cf. Exercise 13). Most measures which arise in practice are σ-finite, which is fortunate since non-σ-finite measures tend to exhibit pathological behavior. The properties of non-σ-finite measures will be explored from time to time in the exercises.

Let us examine a few examples of measures. These examples are of a rather trivial nature, although the first one is of practical importance; the construction of more interesting examples is a task to which we shall turn in the next section.

(i) Let X be any nonempty set, $\mathscr{M} = \mathscr{P}(X)$, and $f\colon X \to [0, \infty]$ an arbitrary function. Then f determines a measure μ on $\mathscr{M}$ by the formula $\mu(E) = \sum_{x \in E} f(x)$. (For the definition of such possibly uncountable sums, see Section 5 of the Prologue.) The reader may verify that μ is semifinite iff $f(x) < \infty$ for every $x \in X$, and μ is σ-finite iff μ is semifinite and $\{x\colon f(x) > 0\}$ is countable. Two special cases are of particular significance: if $f(x) = 1$ for all x, μ is called **counting measure**, and if there exists $x_0 \in X$ such that $f(x_0) = 1$ and $f(x) = 0$ for $x \neq x_0$, then μ is called the **point mass** (or **Dirac measure**) at x_0. (The same names are also applied to the restrictions of these measures to smaller σ-algebras on X.)

(ii) Let X be an uncountable set, and let $\mathscr{M}$ be the σ-algebra of countable or co-countable sets. The function μ on $\mathscr{M}$ defined by $\mu(E) = 0$ if E is countable and $\mu(E) = 1$ if E is co-countable is easily seen to be a measure.

(iii) Let X be an infinite set, $\mathscr{M} = \mathscr{P}(X)$, and define $\mu(E) = 0$ if E is finite, $\mu(E) = \infty$ if E is infinite. Then μ is a finitely additive measure but not a measure.

The basic properties of measures are summarized in the following theorem.

(1.8) Theorem. Let $(X, \mathscr{M}, \mu)$ be a measure space.

(a) (Monotonicity) If $E, F \in \mathscr{M}$ and $E \subset F$, then $\mu(E) \le \mu(F)$.

(b) (Subadditivity) If $\{E_j\}_1^\infty \subset \mathscr{M}$, then $\mu(\bigcup_1^\infty E_j) \le \sum_1^\infty \mu(E_j)$.

(c) (Continuity from below) If $\{E_j\}_1^\infty \subset \mathscr{M}$ and $E_1 \subset E_2 \subset \cdots$, then $\mu(\bigcup_1^\infty E_j) = \lim_{j\to\infty} \mu(E_j)$.

(d) (Continuity from above) If $\{E_j\}_1^\infty \subset \mathscr{M}$, $E_1 \supset E_2 \supset \cdots$, and $\mu(E_n) < \infty$ for some n, then $\mu(\bigcap_1^\infty E_j) = \lim_{j\to\infty} \mu(E_j)$.

Proof. (a) If $E \subset F$, then $\mu(F) = \mu(E) + \mu(F \setminus E) \ge \mu(E)$.

(b) Let $F_1 = E_1$ and $F_k = E_k \setminus (\bigcup_1^{k-1} E_j)$ for $k > 1$. Then the F_k's are disjoint and $\bigcup_1^n F_j = \bigcup_1^n E_j$ for all n. Therefore, by (a),

$$\mu\left(\bigcup_1^\infty E_j\right) = \mu\left(\bigcup_1^\infty F_j\right) = \sum_1^\infty \mu(F_j) \le \sum_1^\infty \mu(E_j).$$

(c) Setting $E_0 = \varnothing$, we have

$$\mu\left(\bigcup_1^\infty E_j\right) = \sum_1^\infty \mu(E_j \setminus E_{j-1}) = \lim_{n\to\infty} \sum_1^n \mu(E_j \setminus E_{j-1}) = \lim_{n\to\infty} \mu(E_n).$$

(d) Let $F_j = E_n \setminus E_j$ for $j > n$: thus $F_{n+1} \subset F_{n+2} \subset \cdots$, $\mu(E_n) = \mu(F_j) + \mu(E_j)$ for $j > n$, and $\bigcup_{n+1}^\infty F_j = E_n \setminus (\bigcap_1^\infty E_j)$. By (c), then,

$$\mu(E_n) = \mu\left(\bigcap_1^\infty E_j\right) + \lim_{j\to\infty} \mu(F_j) = \mu\left(\bigcap_1^\infty E_j\right) + \lim_{j\to\infty} \big(\mu(E_n) - \mu(E_j)\big).$$

Since $\mu(E_n) < \infty$, we may subtract it from both sides to yield the desired result. □

We remark that the condition $\mu(E_n) < \infty$ in (d) is necessary. It is easy to find examples of decreasing sequences $\{E_j\}$ such that $\mu(E_j) = \infty$ for all j but $\bigcap_1^\infty E_j = \varnothing$.

If $(X, \mathcal{M}, \mu)$ is a measure space, a set $E \in \mathcal{M}$ such that $\mu(E) = 0$ is called a **null set**. By subadditivity, any countable union of null sets is null, a fact which we shall use frequently. If a statement about points $x \in X$ is true except for x in some null set, we say that it is true **almost everywhere** (abbreviated **a.e.**) or for **almost every** x. (If more precision is needed, we shall speak of a **μ-null** set, or **μ-almost everywhere.**)

If $\mu(E) = 0$ and $F \subset E$, then $\mu(F) = 0$ by monotonicity provided $F \in \mathcal{M}$, but in general it need not be true that $F \in \mathcal{M}$. (For example, consider the zero measure on the σ-algebra $\{\varnothing, X\}$.) A measure whose domain contains all subsets of null sets is called **complete**. Completeness can sometimes obviate annoying technical points, and it can always be achieved by enlarging the domain of μ, as follows.

(1.9) Theorem. Suppose $(X, \mathcal{M}, \mu)$ is a measure space. Let $\mathcal{N} = \{N \in \mathcal{M}: \mu(N) = 0\}$ and $\overline{\mathcal{M}} = \{E \cup F: E \in \mathcal{M} \text{ and } F \subset N \text{ for some } N \in \mathcal{N}\}$. Then $\overline{\mathcal{M}}$ is a σ-algebra, and there is a unique extension $\bar{\mu}$ of μ to a complete measure on $\overline{\mathcal{M}}$.

Proof. Since $\mathcal{M}$ and $\mathcal{N}$ are closed under countable unions, so is $\overline{\mathcal{M}}$. If $E \cup F \in \overline{\mathcal{M}}$ where $F \subset N \in \mathcal{N}$, we can assume that $E \cap N = \varnothing$ (otherwise, replace F, N by $F \setminus E$, $N \setminus E$). Then $E \cup F = (E \cup N) \cap (N^c \cup F)$, so $(E \cup F)^c = (E \cup N)^c \cup (N \setminus F)$. But $(E \cup N)^c \in \mathcal{M}$ and $N \setminus F \subset N$, so that $(E \cup F)^c \in \overline{\mathcal{M}}$. Thus $\overline{\mathcal{M}}$ is a σ-algebra. If $E \cup F \in \overline{\mathcal{M}}$ $(F \subset N \in \mathcal{N})$, set $\bar{\mu}(E \cup F) = \mu(E)$. This is well defined, since if $E_1 \cup F_1 = E_2 \cup F_2$ $(F_j \subset N_j \in \mathcal{N})$, then $E_1 \subset E_2 \cup N_2$ and so $\mu(E_1) \le \mu(E_2) + \mu(N_2) = \mu(E_2)$, and likewise $\mu(E_2) \le \mu(E_1)$. It is easily verified that $\bar{\mu}$ is a complete measure on $\overline{\mathcal{M}}$, and that $\bar{\mu}$ is the only measure on $\overline{\mathcal{M}}$ which extends μ; details are left to the reader (Exercise 6). □

The measure $\bar{\mu}$ in Theorem (1.9) is called the **completion** of μ, and $\overline{\mathcal{M}}$ is called the **completion** of $\mathcal{M}$ with respect to μ.

EXERCISES

6. Complete the proof of Theorem (1.9).
7. If $\mu_1, \dots, \mu_n$ are measures on $(X, \mathcal{M})$ and $a_1, \dots, a_n \in [0, \infty)$, then $\sum_1^n a_j \mu_j$ is a measure on $(X, \mathcal{M})$.

8. If $(X, \mathcal{M}, \mu)$ is a measure space and $\{E_j\}_1^\infty \subset \mathcal{M}$, then $\mu(\liminf E_j) \le \liminf \mu(E_j)$. Also, $\mu(\limsup E_j) \ge \limsup \mu(E_j)$ provided $\mu(\bigcup_1^\infty E_j) < \infty$.
9. If $(X, \mathcal{M}, \mu)$ is a measure space and $E, F \in \mathcal{M}$, then $\mu(E) + \mu(F) = \mu(E \cup F) + \mu(E \cap F)$.
10. If $(X, \mathcal{M}, \mu)$ is a measure space and $E \in \mathcal{M}$, define $\mu_E(A) = \mu(A \cap E)$ for $A \in \mathcal{M}$. Then μ_E is a measure.
11. A finitely additive measure μ is a measure iff it satisfies the conclusion of Theorem (1.8c). If $\mu(X) < \infty$, μ is a measure iff it satisfies the conclusion of Theorem (1.8d).
12. Let $(X, \mathcal{M}, \mu)$ be a measure space.
    (a) If $E, F \in \mathcal{M}$ and $\mu(E \Delta F) = 0$, then $\mu(E) = \mu(F)$.
    (b) Say that $E \sim F$ iff $\mu(E \Delta F) = 0$: then $\sim$ is an equivalence relation on $\mathcal{M}$.
    (c) If $E, F \in \mathcal{M}$, define $\rho(E, F) = \mu(E \Delta F)$. Then $\rho(E, G) \le \rho(E, F) + \rho(F, G)$ for all $E, F, G \in \mathcal{M}$, and hence ρ defines a metric on the space $\mathcal{M}/\sim$ of equivalence classes.
13. Every σ-finite measure is semifinite.
14. If μ is a semifinite measure and $\mu(E) = \infty$, then for any $C > 0$ there exists $F \subset E$ with $C < \mu(F) < \infty$.
15. If μ is a measure on $(X, \mathcal{M})$, define μ_0 on $\mathcal{M}$ by $\mu_0(E) = \sup\{\mu(F): F \subset E \text{ and } \mu(F) < \infty\}$.
    (a) μ_0 is a semifinite measure. It is called the **semifinite part** of μ.
    (b) If μ is semifinite, then $\mu = \mu_0$. (Use Exercise 14.)
    (c) There is a measure ν on $\mathcal{M}$ (in general, not unique) which assumes only the values 0 and ∞ such that $\mu = \mu_0 + \nu$.
16. Let $(X, \mathcal{M}, \mu)$ be a measure space. A set $E \subset X$ is called **locally measurable** if $E \cap A \in \mathcal{M}$ for all $A \in \mathcal{M}$ such that $\mu(A) < \infty$. Let $\tilde{\mathcal{M}}$ be the collection of all locally measurable sets. Clearly $\mathcal{M} \subset \tilde{\mathcal{M}}$; if $\mathcal{M} = \tilde{\mathcal{M}}$, then μ is called **saturated**.
    (a) If μ is σ-finite, then μ is saturated.
    (b) $\tilde{\mathcal{M}}$ is a σ-algebra.
    (c) Define $\tilde{\mu}$ on $\tilde{\mathcal{M}}$ by $\tilde{\mu}(E) = \mu(E)$ if $E \in \mathcal{M}$, $\tilde{\mu}(E) = \infty$ otherwise. Then $\tilde{\mu}$ is a saturated measure on $\tilde{\mathcal{M}}$, called the **saturation** of μ.
    (d) If μ is complete, so is $\tilde{\mu}$.
    (e) Suppose μ is semifinite. If $E \in \tilde{\mathcal{M}}$, define $\underline{\mu}(E) = \sup\{\mu(A): A \in \mathcal{M} \text{ and } A \subset E\}$. Then $\underline{\mu}$ is a saturated measure on $\tilde{\mathcal{M}}$ which extends μ.
    (f) Let X_1, X_2 be disjoint uncountable sets, $X = X_1 \cup X_2$, and $\mathcal{M}$ the σ-algebra of countable or co-countable sets in X. Define μ on $\mathcal{M}$ by $\mu(E) = \mu_0(E \cap X_1)$ where μ_0 is counting measure on X_1. Then μ is a measure on $\mathcal{M}$, $\tilde{\mathcal{M}} = \mathcal{P}(X)$, and [in the notation of parts (c) and (e)] $\tilde{\mu} \ne \underline{\mu}$.

4. OUTER MEASURES

In this section we develop the tools we shall use to construct measures. If X is a nonempty set, a function $\mu^*\colon \mathscr{P}(X) \to [0, \infty]$ is called an **outer measure** if (i) $\mu^*(\varnothing) = 0$, (ii) $\mu^*(A) \le \mu^*(B)$ whenever $A \subset B$, and (iii) if $\{A_j\}_1^\infty \subset \mathscr{P}(X)$, $\mu^*(\bigcup_1^\infty A_j) \le \sum_1^\infty \mu^*(A_j)$. The name "outer measure" derives from the way in which such objects are usually constructed: one starts with some sort of "proto-measure" on a family $\mathscr{E} \subset \mathscr{P}(X)$ and then approximates arbitrary subsets of X "from the outside" by countable unions of members of $\mathscr{E}$. The precise construction is the following:

(1.10) Proposition. Let $\mathscr{E} \subset \mathscr{P}(X)$ and $\rho\colon \mathscr{E} \to [0, \infty]$ be such that $\varnothing \in \mathscr{E}$, $X \in \mathscr{E}$, and $\rho(\varnothing) = 0$. For any $A \subset X$, let

$$\mu^*(A) = \inf\left\{\sum_1^\infty \rho(E_j)\colon E_j \in \mathscr{E} \text{ and } A \subset \bigcup_1^\infty E_j\right\}.$$

Then μ^* is an outer measure.

Proof. Since $X \in \mathscr{E}$, for any $A \subset X$ there exists $\{E_j\}_1^\infty \subset \mathscr{E}$ such that $A \subset \bigcup_1^\infty E_j$ (take $E_j = X$ for all j), so the definition of μ^* makes sense. Obviously $\mu^*(\varnothing) = 0$ (take $E_j = \varnothing$ for all j) and $\mu^*(A) \le \mu^*(B)$ for $A \subset B$ because the set over which the infimum is taken in the definition of $\mu^*(A)$ includes the corresponding set in the definition of $\mu^*(B)$. If $\{A_j\} \subset \mathscr{P}(X)$ and $\varepsilon > 0$, for each j there exists $\{E_j^k\}_{k=1}^\infty \subset \mathscr{E}$ such that $A_j \subset \bigcup_{k=1}^\infty E_j^k$ and $\sum_{k=1}^\infty \rho(E_j^k) \le \mu^*(A_j) + \varepsilon 2^{-j}$. But then if $A = \bigcup_1^\infty A_j$ we have $A \subset \bigcup_{j,k=1}^\infty E_j^k$ and $\sum_{j,k} \rho(E_j^k) \le \sum_j \mu^*(A_j) + \varepsilon$, whence $\mu^*(A) \le \sum_j \mu^*(A_j) + \varepsilon$. Since ε is arbitrary, we are done. □

If μ^* is an outer measure on X, a set $A \subset X$ is called **μ^*-measurable** if

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \quad \text{for all } E \subset X.$$

Of course, the inequality $\mu^*(E) \le \mu^*(E \cap A) + \mu^*(E \cap A^c)$ holds for any A and E, so to prove that A is μ^*-measurable it suffices to prove the reverse inequality. The latter is trivial if $\mu^*(E) = \infty$, so we see that A is μ^*-measurable iff

$$\mu^*(E) \ge \mu^*(E \cap A) + \mu^*(E \cap A^c) \quad \text{for all } E \subset X \text{ such that } \mu^*(E) < \infty.$$

The notion of μ^*-measurability is perhaps not very intuitive at first, but it is extremely useful because of the following theorem.

(1.11) Carathéodory's Theorem. If μ^* is an outer measure on X, the collection $\mathscr{M}$ of μ^*-measurable sets is a σ-algebra, and the restriction of μ^* to $\mathscr{M}$ is a complete measure.

Proof. $\mathscr{M}$ is closed under complements since the definition of μ^*-measurability of A is symmetric in A and A^c. Next, if $A, B \in \mathscr{M}$ and $E \subset X$, by subadditivity we have

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c)\\ &= \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c)\\ &\quad + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c)\\ &\ge \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c).\end{aligned}$$

It follows that $A \cup B \in \mathscr{M}$, so $\mathscr{M}$ is an algebra. Moreover, if $A, B \in \mathscr{M}$ and $A \cap B = \varnothing$, then

$$\mu^*(A \cup B) = \mu^*((A \cup B) \cap A) + \mu^*((A \cup B) \cap A^c) = \mu^*(A) + \mu^*(B),$$

so μ^* is finitely additive on $\mathscr{M}$. To show that $\mathscr{M}$ is a σ-algebra it will suffice to show that $\mathscr{M}$ is closed under countable disjoint unions. If $\{A_j\}_1^\infty$ is a sequence of disjoint sets in $\mathscr{M}$, let $B_n = \bigcup_1^n A_j$ and $B = \bigcup_1^\infty A_j$. Then for any $E \subset X$,

$$\begin{aligned}\mu^*(E \cap B_n) &= \mu^*(E \cap B_n \cap A_n) + \mu^*(E \cap B_n \cap A_n^c)\\ &= \mu^*(E \cap A_n) + \mu^*(E \cap B_{n-1}),\end{aligned}$$

so a simple induction shows that $\mu^*(E \cap B_n) = \sum_1^n \mu^*(E \cap A_j)$. Therefore,

$$\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \ge \sum_1^n \mu^*(E \cap A_j) + \mu^*(E \cap B^c),$$

and letting $n \to \infty$ we obtain

$$\begin{aligned}\mu^*(E) &\ge \sum_1^\infty \mu^*(E \cap A_j) + \mu^*(E \cap B^c) \ge \mu^*\left(\bigcup_1^\infty (E \cap A_j)\right) + \mu^*(E \cap B^c)\\ &= \mu^*(E \cap B) + \mu^*(E \cap B^c) \ge \mu^*(E).\end{aligned}$$

All the inequalities in this last calculation are thus equalities. It follows that $B \in \mathscr{M}$ and—taking $E = B$—that $\mu^*(B) = \sum_1^\infty \mu^*(A_j)$, so μ^* is countably additive on $\mathscr{M}$. Finally, if $\mu^*(A) = 0$, for any $E \subset X$ we have

$$\mu^*(E) \le \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E \cap A^c) \le \mu^*(E),$$

so that $A \in \mathscr{M}$. Therefore $\mu^* | \mathscr{M}$ is a complete measure. □

We shall now apply Carathéodory's theorem to the problem of extending measures from algebras to σ-algebras. If $\mathscr{A} \subset \mathscr{P}(X)$ is an algebra, a function $\mu\colon \mathscr{A} \to [0, \infty]$ will be called a **premeasure** (i) if $\mu(\varnothing) = 0$, (ii) if $\{A_j\}_1^\infty \subset \mathscr{A}$ is a sequence of disjoint sets such that $\bigcup_1^\infty A_j \in \mathscr{A}$, then $\mu(\bigcup_1^\infty A_j) = \sum_1^\infty \mu(A_j)$. (In particular, a premeasure is finitely additive, for one can take $A_j = \varnothing$ for j large.) The notions of finite and σ-finite premeasures are defined just as for measures. If μ is a premeasure on $\mathscr{A} \subset \mathscr{P}(X)$, it induces an outer measure on X in accordance with Proposition (1.10), namely

$$\mu^*(E) = \inf\left\{\sum_1^\infty \mu(A_j)\colon A_j \in \mathscr{A}, E \subset \bigcup_1^\infty A_j\right\}. \tag{1.12}$$

(1.13) Proposition. If μ is a premeasure on $\mathscr{A}$ and μ^* is defined by (1.12), then (a) $\mu^*|\mathscr{A} = \mu$ and (b) every set in $\mathscr{A}$ is μ^*-measurable.

Proof. (a) If $E \in \mathscr{A}$ and $E \subset \bigcup_1^\infty A_j$ with $A_j \in \mathscr{A}$, let $B_n = E \cap (A_n \setminus (\bigcup_1^{n-1} A_j))$. Then the B_n's are disjoint members of $\mathscr{A}$ whose union is E, so $\mu(E) = \sum_1^\infty \mu(B_j) \le \sum_1^\infty \mu(A_j)$. It follows that $\mu(E) \le \mu^*(E)$, and the reverse inequality is obvious since $E \subset \bigcup_1^\infty A_j$ where $A_1 = E$ and $A_j = \varnothing$ for $j > 1$.

(b) If $A \in \mathscr{A}$, $E \subset X$, and $\varepsilon > 0$, there is a sequence $\{B_j\}_1^\infty \subset \mathscr{A}$ with $E \subset \bigcup_1^\infty B_j$ and $\sum_1^\infty \mu(B_j) \le \mu^*(E) + \varepsilon$. Since μ is additive on $\mathscr{A}$,

$$\mu^*(E) + \varepsilon \ge \sum_1^\infty \mu(B_j \cap A) + \sum_1^\infty \mu(B_j \cap A^c)$$

$$\ge \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

Since ε is arbitrary, A is μ^*-measurable. □

(1.14) Theorem. Let $\mathscr{A} \subset \mathscr{P}(X)$ be an algebra, μ a premeasure on $\mathscr{A}$, and $\mathscr{M}$ the σ-algebra generated by $\mathscr{A}$. There exists a measure $\bar{\mu}$ on $\mathscr{M}$ whose restriction to $\mathscr{A}$ is μ—namely, $\bar{\mu} = \mu^*|\mathscr{M}$ where μ^* is given by (1.12). If ν is another such measure on $\mathscr{M}$, then $\nu(E) \le \bar{\mu}(E)$ for all $E \in \mathscr{M}$, with equality when $\bar{\mu}(E) < \infty$. If μ is σ-finite, then $\bar{\mu}$ is the unique extension of μ to a measure on $\mathscr{M}$.

Proof. The first assertion follows from Carathéodory's theorem and Proposition (1.13) since the σ-algebra of μ^*-measurable sets includes $\mathscr{A}$ and hence $\mathscr{M}$. If $E \in \mathscr{M}$ and $E \subset \bigcup_1^\infty A_j$ where $A_j \in \mathscr{A}$, then $\nu(E) \le \sum_1^\infty \nu(A_j) = \sum_1^\infty \mu(A_j)$, whence $\nu(E) \le \bar{\mu}(E)$. Also, observe that if $A = \bigcup_1^\infty A_j$, we have

$$\nu(A) = \lim_{n\to\infty} \nu\left(\bigcup_1^n A_j\right) = \lim_{n\to\infty} \mu\left(\bigcup_1^n A_j\right) = \bar{\mu}(A).$$

If $\bar{\mu}(E) < \infty$, we can choose the A_j's so that $\bar{\mu}(A) < \bar{\mu}(E) + \varepsilon$, hence $\bar{\mu}(A \setminus E) < \varepsilon$, and

$$\bar{\mu}(E) \le \bar{\mu}(A) = \nu(A) = \nu(E) + \nu(A \setminus E)$$

$$\le \nu(E) + \bar{\mu}(A \setminus E) \le \nu(E) + \varepsilon.$$

Since ε is arbitrary, $\bar{\mu}(E) = \nu(E)$. Finally, if $X = \bigcup_1^\infty A_j$ with $\mu(A_j) < \infty$, we can assume the A_j's are disjoint. Then for any $E \in \mathscr{M}$,

$$\bar{\mu}(E) = \sum_1^\infty \bar{\mu}(E \cap A_j) = \sum_1^\infty \nu(E \cap A_j) = \nu(E),$$

so $\bar{\mu} = \nu$. □

The proof of this theorem yields more than the statement: indeed, μ may be extended to a measure on the algebra $\mathscr{M}^*$ of all μ^*-measurable sets. The relation between $\mathscr{M}$ and $\mathscr{M}^*$ is explored in Exercises 20b and 22.

EXERCISES

17. If μ^* is an outer measure and $\{A_j\}_1^\infty$ is a sequence of disjoint μ^*-measurable sets, then $\mu^*(E \cap (\bigcup_1^\infty A_j)) = \sum_1^\infty \mu^*(E \cap A_j)$ for any $E \subset X$.

18. Let $\mathscr{A} \subset \mathscr{P}(X)$ be an algebra, $\mathscr{A}_\sigma$ the collection of countable unions of sets in $\mathscr{A}$, and $\mathscr{A}_{\sigma\delta}$ the collection of countable intersections of sets in $\mathscr{A}_\sigma$. Let μ be a premeasure on $\mathscr{A}$ and μ^* the induced outer measure.

(a) For any $E \subset X$ and $\varepsilon > 0$ there exists $A \in \mathscr{A}_\sigma$ with $E \subset A$ and $\mu^*(A) \le \mu^*(E) + \varepsilon$.

(b) If $\mu^*(E) < \infty$, then E is μ^*-measurable iff there exists $B \in \mathscr{A}_{\sigma\delta}$ with $E \subset B$ and $\mu^*(B \setminus E) = 0$.

(c) If μ is σ-finite, the restriction $\mu^*(E) < \infty$ in (b) is superfluous.

19. Let μ^* be an outer measure on X induced from a premeasure, such that $\mu^*(X) < \infty$. If $E \subset X$, define the **inner measure** of E to be $\mu_*(E) = \mu^*(X) - \mu^*(E^c)$. Then E is μ^*-measurable iff $\mu^*(E) = \mu_*(E)$. (Use Exercise 18a).

20. Let μ^* be an outer measure on X, $\mathscr{M}^*$ the σ-algebra of μ^*-measurable sets, $\bar{\mu} = \mu^*|\mathscr{M}^*$, and μ^+ the outer measure induced by $\bar{\mu}$ according to (1.12).

(a) If $E \subset X$, we have $\mu^*(E) \le \mu^+(E)$, with equality iff there exists $A \in \mathscr{M}^*$ with $A \supset E$ and $\mu^*(A) = \mu^*(E)$.

(b) If μ^* is induced from a premeasure, then $\mu^* = \mu^+$. (Use Exercise 18a.)

(c) Construct a μ^* on $X = \{0, 1\}$ such that $\mu^* \ne \mu^+$.

21. Let μ^* be an outer measure on X and $\bar{\mu}$ the restriction of μ^* to the μ^*-measurable sets. Then $\bar{\mu}$ is saturated. (Cf. Exercise 16.)

22. Let $(X, \mathcal{M}, \mu)$ be a measure space, μ^* the outer measure induced by μ according to (1.12), $\mathcal{M}^*$ the σ-algebra of μ^*-measurable sets, and $\tilde{\mu} = \mu^*|\mathcal{M}^*$.

(a) If μ is σ-finite, then $\tilde{\mu}$ is the completion of μ. (Use Exercise 18.)

(b) In the general case, $\tilde{\mu}$ is the saturation of the completion of μ. (Cf. Exercises 16 and 21.)

23. Let $\mathcal{A}$ be the collection of finite unions of sets of the form $(a, b] \cap \mathbf{Q}$ $(-\infty \le a \le b \le \infty)$.

(a) $\mathcal{A}$ is an algebra on $\mathbf{Q}$. [Use Proposition (1.7).]

(b) The σ-algebra generated by $\mathcal{A}$ is $\mathcal{P}(\mathbf{Q})$.

(c) Define μ on $\mathcal{A}$ by $\mu(\varnothing) = 0$ and $\mu(A) = \infty$ for $A \neq \varnothing$. Then μ is a premeasure on $\mathcal{A}$, and there is more than one measure on $\mathcal{P}(\mathbf{Q})$ whose restriction to $\mathcal{A}$ is μ.

24. Let $(X, \mathcal{M}, \mu)$ be a measure space such that $\mu(X) = 1$, and let μ^* be the outer measure induced by μ. Suppose $E \subset X$ satisfies $\mu^*(E) = 1$.

(a) If $A, B \in \mathcal{M}$ and $A \cap E = B \cap E$, then $\mu(A) = \mu(B)$.

(b) Let $\mathcal{M}_E = \{A \cap E: A \in \mathcal{M}\}$. According to (a) the function ν on $\mathcal{M}_E$ given by $\nu(A \cap E) = \mu(A)$ is well defined. $\mathcal{M}_E$ is a σ-algebra on E and ν is a measure on $\mathcal{M}_E$.

5. BOREL MEASURES ON R

We now apply the general theory of the previous section to construct Borel measures on $\mathbf{R}$—that is, measures whose domain is $\mathcal{B}_\mathbf{R}$. In this section we shall refer to left-open, right-closed intervals in $\mathbf{R}$ (i.e., sets of the form $(a, b]$ or (a, ∞) or $\varnothing$) as **h-intervals** (h for "half-open"). Clearly the intersection of two h-intervals is an h-interval, and the complement of an h-interval is an h-interval or the disjoint union of two h-intervals. By Proposition (1.7), the collection $\mathcal{A}$ of finite disjoint unions of h-intervals is an algebra, and by Proposition (1.2), the σ-algebra generated by $\mathcal{A}$ is the Borel σ-algebra $\mathcal{B}_\mathbf{R}$.

If F: $\mathbf{R} \to \mathbf{R}$ is an increasing function, then F has right- and left-hand limits at each point:

$$F(a^+) = \lim_{x \to a^+} F(x) = \inf_{x > a} F(x),$$

$$F(a^-) = \lim_{x \to a^-} F(x) = \sup_{x > a} F(x).$$

Moreover, the limiting values $F(\infty) = \sup_{x \in \mathbf{R}} F(x)$ and $F(-\infty) =$ $\inf_{x \in \mathbf{R}} F(x)$ exist (possibly equal to $\pm\infty$). We say that F is **right continuous** if $F(a) = F(a^+)$ for all $a \in \mathbf{R}$.

(1.15) Proposition. Let F: $\mathbf{R} \to \mathbf{R}$ be increasing and right continuous. If $(a_j, b_j]$ $(j = 1, \ldots, n)$ are disjoint h-intervals, let

$$\mu\left(\bigcup_1^n (a_j, b_j]\right) = \sum_1^n \left[F(b_j) - F(a_j)\right],$$

and let $\mu(\varnothing) = 0$. Then μ is a premeasure on the algebra $\mathcal{A}$.

Proof. First we must check that μ is well defined, since elements of $\mathcal{A}$ can be represented in more than one way as disjoint unions of h-intervals. If $\{(a_j, b_j]\}_1^n$ are disjoint and $\bigcup_1^n (a_j, b_j] = (a, b]$, then, after perhaps relabeling the index j, we must have $a = a_1 < b_1 = a_2 < b_2 = \cdots < b_n = b$, so

$$\sum_1^n \left[F(b_j) - F(a_j)\right] = F(b) - F(a).$$

More generally, if $\{I_i\}_1^n$ and $\{J_j\}_1^m$ are finite sequences of disjoint h-intervals such that $\bigcup_1^n I_i = \bigcup_1^m J_j$, this reasoning shows that

$$\sum_1^n \mu(I_i) = \sum_1^n \sum_1^m \mu(I_i \cap J_j) = \sum_1^m \mu(J_j).$$

Thus μ is well defined, and it is finitely additive by definition.

It remains to show that if $\{I_j\}_1^\infty$ is a sequence of disjoint h-intervals with $\bigcup_1^\infty I_j \in \mathcal{A}$ then $\mu(\bigcup_1^\infty I_j) = \sum_1^\infty \mu(I_j)$. By finite additivity we may assume that $\bigcup_1^\infty I_j$ is an h-interval $I = (a, b]$, and we have

$$\mu(I) = \mu\left(\bigcup_1^n I_j\right) + \mu\left(I \setminus \bigcup_1^n I_j\right) \ge \mu\left(\bigcup_1^n I_j\right) = \sum_1^n \mu(I_j).$$

Letting $n \to \infty$ we obtain $\mu(I) \ge \sum_1^\infty \mu(I_j)$. To prove the reverse inequality, let us suppose first that $-\infty < a < b < \infty$, and let us fix $\varepsilon > 0$. Since F is right continuous, there exists $\delta > 0$ such that $F(a + \delta) - F(a) < \varepsilon$, and if $I_j = (a_j, b_j]$, for each j there exists $\delta_j > 0$ such that $F(b_j + \delta_j) - F(b_j) < \varepsilon 2^{-j}$. The open intervals $(a_j, b_j + \delta_j)$ cover the compact set $[a + \delta, b]$, so there is a finite subcover. By discarding any $(a_j, b_j + \delta_j)$ which is contained in a larger one and

relabeling the index j, we may assume that (i) the intervals $(a_1, b_1 + \delta_1), \dots, (a_N, b_N + \delta_N)$ cover $[a + \delta, b]$, (ii) $a_1 < a_2 < \cdots < a_N$, (iii) $b_j + \delta_j \in (a_{j+1}, b_{j+1} + \delta_{j+1})$ for $j = 1, \dots, N-1$. But then

$$\begin{aligned}
\mu(I) &\le F(b) - F(a+\delta) + \varepsilon \\
&\le F(b_N + \delta_N) - F(a_1) + \varepsilon \\
&= F(b_N + \delta_N) - F(a_N) + \sum_1^{N-1} \big[F(a_{j+1}) - F(a_j)\big] + \varepsilon \\
&\le F(b_N + \delta_N) - F(a_N) + \sum_1^{N-1} \big[F(b_j + \delta_j) - F(a_j)\big] + \varepsilon \\
&\le \sum_1^{\infty} \mu(I_j) + 2\varepsilon.
\end{aligned}$$

Since ε is arbitrary, we are done in the case $-\infty < a < b < \infty$. If $a = -\infty$, the same reasoning gives $F(b) - F(-M) \le \sum_1^\infty \mu(I_j) + 2\varepsilon$ for any $M < \infty$, while if $b = \infty$ we likewise obtain $F(M) - F(a) \le \sum_1^\infty \mu(I_j) + 2\varepsilon$. The desired result then follows by letting $\varepsilon \to 0$ and $M \to \infty$. □

(1.16) Theorem. If F: $\mathbf{R} \to \mathbf{R}$ is any increasing, right continuous function, there is a unique measure μ_F on $\mathscr{B}_\mathbf{R}$ such that $\mu_F((a, b]) = F(b) - F(a)$ for all $a, b \in \mathbf{R}$. If G is another such function, we have $\mu_F = \mu_G$ iff $F - G$ is constant. Conversely, if μ is a measure on $\mathscr{B}_\mathbf{R}$ which is finite on all bounded Borel sets and we define

$$F(x) = \mu((0, x]) \text{ if } x > 0, \qquad F(0) = 0, \qquad F(x) = -\mu((x, 0]) \text{ if } x < 0,$$

then F is increasing and right continuous, and $\mu = \mu_F$.

Proof. Each F induces a premeasure on $\mathscr{A}$ by Proposition (1.15). It is clear that F and G induce the same premeasure iff $F - G$ is constant, and that these premeasures are σ-finite (since $\mathbf{R} = \bigcup_{-\infty}^{\infty}(j, j+1]$). The first two assertions therefore follow from Theorem (1.14). As for the last one, the monotonicity of μ implies the monotonicity of F, and the continuity of μ from above and below implies the right continuity of F for $x \ge 0$ and $x < 0$. It is evident that $\mu = \mu_F$ on $\mathscr{A}$, and hence $\mu = \mu_F$ on $\mathscr{B}_\mathbf{R}$ by the uniqueness assertion of Theorem (1.14). □

Several remarks are in order. First, this theory could equally well be developed by using half-open intervals of the form $[a, b)$ and left continuous functions F. Second, if μ is a finite measure on $\mathscr{B}_\mathbf{R}$, then $\mu = \mu_F$ where $F(x) = \mu((-\infty, x])$. This F is called the **cumulative distribution function** of μ.

Third, the theory of Section 4 gives, for each increasing, right continuous F, a complete measure $\bar{\mu}_F$ whose domain is almost always bigger than $\mathscr{B}_\mathbf{R}$. In fact, $\bar{\mu}_F$ is just the completion of μ_F (cf. Exercises 20b and 22). We shall usually denote this complete measure also by μ_F; it is called the **Lebesgue–Stieltjes measure** associated to F.

Lebesgue–Stieltjes measures enjoy some useful regularity properties which we now investigate. In this discussion we fix an increasing, right continuous function F; we denote by μ the complete Lebesgue–Stieltjes measure associated to F and by $\mathscr{M}$ the domain of μ. Thus, for any $E \in \mathscr{M}$,

$$\begin{aligned}
\mu(E) &= \inf\Big\{\sum_1^\infty \big[F(b_j) - F(a_j)\big] : E \subset \bigcup_1^\infty (a_j, b_j]\Big\} \\
&= \inf\Big\{\sum_1^\infty \mu\big((a_j, b_j]\big) : E \subset \bigcup_1^\infty (a_j, b_j]\Big\}.
\end{aligned}$$

We first observe that in the second formula for $\mu(E)$ we can replace h-intervals by open intervals:

(1.17) Lemma. For any $E \in \mathscr{M}$,

$$\mu(E) = \inf\Big\{\sum_1^\infty \mu\big((a_j, b_j)\big) : E \subset \bigcup_1^\infty (a_j, b_j)\Big\}.$$

Proof. Let us denote the quantity on the right by $\nu(E)$. If $E \subset \bigcup_1^\infty (a_j, b_j)$, let $l_j = b_j - a_j$ and $I_{jk} = (b_j - l_j 2^{1-k}, b_j - l_j 2^{-k}]$ $(k \in \mathbf{N})$. Then $(a_j, b_j) = \bigcup_{k=1}^\infty I_{jk}$, so $E \subset \bigcup_{j,k=1}^\infty I_{jk}$ and

$$\sum_1^\infty \mu\big((a_j, b_j)\big) = \sum_{j,k=1}^\infty \mu(I_{jk}) \ge \mu(E).$$

Thus $\nu(E) \ge \mu(E)$. On the other hand, given $\varepsilon > 0$ there exists $\{(a_j, b_j]\}_1^\infty$ with $E \subset \bigcup_1^\infty (a_j, b_j]$ and $\sum_1^\infty \mu((a_j, b_j]) \le \mu(E) + \varepsilon$, and for each j there exists $\delta_j > 0$ such that $F(b_j + \delta_j) - F(b_j) < \varepsilon 2^{-j}$. Then $E \subset \bigcup_1^\infty (a_j, b_j + \delta_j)$ and

$$\sum_1^\infty \mu\big((a_j, b_j + \delta_j)\big) \le \sum_1^\infty \mu\big((a_j, b_j]\big) + \varepsilon \le \mu(E) + 2\varepsilon,$$

so that $\nu(E) \le \mu(E)$. □

(1.18) Theorem. If $E \in \mathscr{M}$, then

$$\begin{aligned}
\mu(E) &= \inf\{\mu(U) : U \supset E \text{ and } U \text{ is open}\} \\
&= \sup\{\mu(K) : K \subset E \text{ and } K \text{ is compact}\}.
\end{aligned}$$

Proof. If U is open and $U \supset E$, then $\mu(U) \geq \mu(E)$; on the other hand, U is a countable union of open intervals (a_j, b_j), and $\mu(U) \leq \sum_1^\infty \mu((a_j, b_j))$. The first equation therefore follows from Lemma (1.17). For the second one, suppose first that E is bounded. If $E = \bar{E}$, E is compact and the equation is obvious. Otherwise, given $\varepsilon > 0$ we can choose an open $U \supset \bar{E} \setminus E$ such that $\mu(U) \leq \mu(\bar{E} \setminus E) + \varepsilon$. Let $K = \bar{E} \setminus U$. Then K is compact, $K \subset E$, and

$$\begin{aligned} \mu(K) &= \mu(E) - \mu(E \cap U) \\ &= \mu(E) - [\mu(U) - \mu(U \setminus E)] \\ &\geq \mu(E) - \mu(U) + \mu(\bar{E} \setminus E) \\ &\geq \mu(E) - \varepsilon. \end{aligned}$$

If E is unbounded, let $E_j = E \cap (j, j+1]$. By the preceding argument, for any $\varepsilon > 0$ there exist compact $K_j \subset E_j$ with $\mu(K_j) \geq \mu(E_j) - \varepsilon 2^{-j}$. Let $H_n = \bigcup_{-n}^{n} K_j$. Then H_n is compact, $H_n \subset E$, and $\mu(H_n) \geq \mu(\bigcup_{-n}^{n} E_j) - \varepsilon$. Since $\mu(E) = \lim_{n \to \infty} \mu(\bigcup_{-n}^{n} E_j)$, the result follows. □

(1.19) Theorem. If $E \subset \mathbf{R}$, the following are equivalent:

(a) $E \in \mathcal{M}$.

(b) $E = V \setminus N_1$ where V is a G_δ set and $\mu(N_1) = 0$.

(c) $E = H \cup N_2$ where H is an F_σ set and $\mu(N_2) = 0$.

Proof. Obviously (b) and (c) imply (a) (keep in mind that μ is complete). Suppose $E \in \mathcal{M}$ and E is bounded. By Theorem (1.18), for $j \in \mathbf{N}$ we can choose an open $U_j \supset E$ and a compact $K_j \subset E$ such that

$$\mu(U_j) - 2^{-j} \leq \mu(E) \leq \mu(K_j) + 2^{-j},$$

Let $V = \bigcap_1^\infty U_j$, $H = \bigcup_1^\infty K_j$. Then $\mu(V) = \mu(E) = \mu(H) < \infty$, so $\mu(V \setminus E) = \mu(E \setminus H) = 0$. The result is thus proved when E is bounded; the unbounded case is left to the reader (Exercise 25). □

The significance of Theorem (1.19) is that all Borel sets (or more generally all sets in $\mathcal{M}$) are of a reasonably simple form modulo sets of measure zero. This contrasts markedly with the machinations necessary to construct the Borel sets from the open sets when null sets are not excepted; see Proposition (1.23) below.

(1.20) Proposition. If $\mu(E) < \infty$, then for every $\varepsilon > 0$ there is a set A which is a finite union of open intervals such that $\mu(E \Delta A) < \varepsilon$.

The proof is left to the reader (Exercise 26).

We now examine the most important measure on $\mathbf{R}$, namely **Lebesgue measure**: this is the complete measure μ_F associated to the function $F(x) = x$ (so that the measure of an interval is simply its length). We shall denote it by m. The domain of m is called the class of **Lebesgue measurable** sets, and we shall denote it by $\mathcal{L}$. We shall also refer to the restriction of m to $\mathcal{B}_\mathbf{R}$ as Lebesgue measure.

(1.21) Theorem. If $E \subset \mathbf{R}$ and $s, r \in \mathbf{R}$, let $E + s = \{x + s: s \in E\}$ and $rE = \{rx: x \in E\}$. If $E \in \mathcal{L}$, then $E + s \in \mathcal{L}$ and $rE \in \mathcal{L}$ for all s and r; moreover, $m(E + s) = m(E)$ and $m(rE) = |r| m(E)$.

Proof. Since the collection of open intervals is invariant under translations and dilations, the same is true of $\mathcal{B}_\mathbf{R}$. For $E \in \mathcal{B}_\mathbf{R}$ let $m_s(E) = m(E + s)$ and $m^r(E) = m(rE)$. Then m_s and m^r clearly agree with m and $|r|m$ on finite unions of intervals, hence on $\mathcal{B}_\mathbf{R}$ by Theorem (1.14). In particular, if $E \in \mathcal{B}_\mathbf{R}$ and $m(E) = 0$, then $m(E + s) = m(rE) = 0$. It follows that $\mathcal{L}$ is preserved by translations and dilations and that $m(E + s) = m(E)$ and $m(rE) = |r| m(E)$ for all $E \in \mathcal{L}$. □

It is possible (although of little practical use) to extend Lebesgue measure to σ-algebras larger than $\mathcal{L}$ while maintaining translation invariance and countable additivity. However, it is not possible to extend it thus to $\mathcal{P}(\mathbf{R})$. In fact, if N is the set constructed in Section 1, the argument given there shows that if μ is a translation-invariant measure on a translation-invariant σ-algebra $\mathcal{M}$ containing $[0, 1)$ such that $0 < \mu([0, 1)) < \infty$, then $N \notin \mathcal{M}$.

Each singleton set in $\mathbf{R}$ has Lebesgue measure zero, and hence so does every countable set. [In particular, $m(\mathbf{Q}) = 0$.] However, there are also Lebesgue null sets having the cardinality of the continuum. We now present the standard example, the Cantor set, which is also of interest for other reasons.

Each $x \in [0, 1]$ has a base-3 decimal expansion $x = \sum_1^\infty a_j 3^{-j}$ where $a_j = 0$, 1, or 2. This expansion is unique unless x is of the form $p3^{-k}$ for some integers p, k, in which case x has two expansions: one with $a_j = 0$ for $j > k$ and one with $a_j = 2$ for $j > k$. Assuming p is nondivisible by 3, one of these expansions will have $a_k = 1$ and the other will have $a_k = 0$ or 2. If we agree always to use the latter expansion, we see that $a_1 = 1$ iff $1/3 < x < 2/3$, $a_1 \neq 1$ and $a_2 = 1$ iff $1/9 < x < 2/9$ or $7/9 < x < 8/9$, and so forth. It will also be useful to observe that if $x = \sum a_j 3^{-j}$ and $y = \sum b_j 3^{-j}$, then $x < y$ iff for some n we have $a_j = b_j$ for $j < n$ and $a_n < b_n$.

The **Cantor set** C is the set of all $x \in [0, 1]$ which have a base-3 expansion $x = \sum a_j 3^{-j}$ with $a_j \neq 1$ for all j. Thus C is obtained from $[0, 1]$ by removing the open middle third $(1/3.2/3)$, then removing the open middle thirds $(1/9, 2/9)$ and $(7/9, 8/9)$ of the two remaining intervals, and so forth. C has several remarkable properties:

(1.22) Proposition. Let C be the Cantor set.

(a) C is compact.

(b) If $x, y \in C$ and $x < y$, there exists $z \notin C$ such that $x < z < y$; thus C is totally disconnected and nowhere dense.

(c) C has no isolated points.

(d) $m(C) = 0$.

(e) C can be mapped onto $[0,1]$; hence $\text{card}(C) = \mathfrak{c}$.

Proof. The proofs of (a), (b), and (c) are left to the reader (Exercise 27). As for (d), C is obtained from $[0,1]$ by removing one interval of length $1/3$, two intervals of length $1/9$, and so forth. Thus

$$m(C) = 1 - \sum_0^\infty \frac{2^j}{3^{j+1}} = 1 - \frac{1}{3} \cdot \frac{1}{1-(2/3)} = 0.$$

(e) can be proved as follows: if $x \in C$, then $x = \Sigma a_j 3^{-j}$ where $a_j = 0$ or 2 for all j. Let $b_j = a_j/2$ and $f(x) = \Sigma b_j 2^{-j}$. This last series is the base-2 expansion of a number in $[0,1]$, and clearly any element of $[0,1]$ can be obtained in this way, so f maps C onto $[0,1]$. □

Let us examine the map f in the preceding proof more closely. One readily sees that if $x, y \in C$ and $x < y$, then $f(x) < f(y)$ unless x and y are the endpoints of one of the intervals removed from $[0,1]$ to obtain C. In this case $f(x) = p2^{-k}$ for some integers p, k and $f(x)$ and $f(y)$ are the two base-2 expansions of this number. We can therefore extend f to a map from $[0,1]$ to itself by defining its value on each interval missing from C to be its value at the endpoints. This extended f is still increasing, and since its range is all of $[0,1]$ it cannot have any jump discontinuities; hence it is continuous. f is called the **Cantor function** or **Cantor–Lebesgue function**.

Not every Lebesgue measurable set is a Borel set. One can display examples of sets in $\mathscr{L} \setminus \mathscr{B}_{\mathbf{R}}$ by using the Cantor function; see Exercise 2.9. Alternatively, one can observe that since every subset of the Cantor set is Lebesgue measurable, we have $\text{card}(\mathscr{L}) = \text{card}(\mathscr{P}(\mathbf{R})) > \mathfrak{c}$, whereas $\text{card}(\mathscr{B}_{\mathbf{R}}) = \mathfrak{c}$. [The latter fact follows from Proposition (1.23).]

EXERCISES

25. Complete the proof of Theorem (1.19).

26. Prove Proposition 1.20. [Use Theorem (1.18).]

27. Prove Proposition (1.22a–c).

28. Let F be increasing and right continuous, and let μ_F be the associated measure. Then $\mu_F(\{a\}) = F(a) - F(a^-)$, $\mu_F([a,b)) = F(b^-) - F(a^-)$, $\mu_F([a,b]) = F(b) - F(a^-)$, and $\mu_F((a,b)) = F(b^-) - F(a)$.

29. Let E be a Lebesgue measurable set.

(a) If $E \subset N$ where N is the nonmeasurable set described in Section 1, then $m(E) = 0$.

(b) If $m(E) > 0$, then E contains a nonmeasurable set. (It suffices to assume $E \subset [0,1)$. In the notation of Section 1, $E = \bigcup_{r \in R} E \cap N_r$.)

30. If $E \in \mathscr{L}$ and $m(E) > 0$, for any $\alpha < 1$ there is an open interval I such that $m(E \cap I) > \alpha m(I)$.

31. If $E \in \mathscr{L}$ and $m(E) > 0$, the set $E - E = \{x - y\colon x, y \in E\}$ contains an interval centered at 0. [If I is as in Exercise 30 with $\alpha > 3/4$, then $E - E$ contains $(-1/2m(I), 1/2m(I))$.]

6. NOTES AND REFERENCES

The history of measure theory is intimately connected with the history of integration theory, comments on which will be made in Section 2.7.

Section 1. The Banach–Tarski paradox appeared first in [8]; an elementary exposition can be found in Stromberg [104].

Section 2. Our characterization of the σ-algebra $\mathscr{M}(\mathscr{E})$ generated by a family $\mathscr{E} \subset \mathscr{P}(X)$ is nonconstructive, and one might ask how to obtain $\mathscr{M}(\mathscr{E})$ explicitly from $\mathscr{E}$. The answer is rather complicated. One can begin as follows: let $\mathscr{E}_1 = \mathscr{E} \cup \{E^c\colon E \in \mathscr{E}\}$, and for $j > 1$ define $\mathscr{E}_j$ to be the collection of all sets which are countable unions of sets in $\mathscr{E}_{j-1}$ or complements of such. Let $\mathscr{E}_\omega = \bigcup_1^\infty \mathscr{E}_j$: is $\mathscr{E}_\omega = \mathscr{M}(\mathscr{E})$? In general, no. $\mathscr{E}_\omega$ is closed under complements, but if $E_j \in \mathscr{E}_j \setminus \mathscr{E}_{j-1}$ for each j, there is no reason for $\bigcup_1^\infty E_j$ to be in $\mathscr{E}_\omega$. So one must start all over again. More precisely, one must define $\mathscr{E}_\alpha$ for every countable ordinal α by transfinite induction: if α has an immediate predecessor β, $\mathscr{E}_\alpha$ is the collection of sets which are countable unions of sets in $\mathscr{E}_\beta$ or complements of such; otherwise, $\mathscr{E}_\alpha = \bigcup_{\beta < \alpha} \mathscr{E}_\beta$. Then:

(1.23) Proposition. $\mathscr{M}(\mathscr{E}) = \bigcup_{\alpha \in \Omega} \mathscr{E}_\alpha$, where Ω is the set of countable ordinals.

Proof. Transfinite induction shows that $\mathscr{E}_\alpha \subset \mathscr{M}(\mathscr{E})$ for all $\alpha \in \Omega$, and hence $\bigcup_{\alpha \in \Omega} \mathscr{E}_\alpha \subset \mathscr{M}(\mathscr{E})$. The reverse inclusion follows from the fact that any sequence in Ω has a supremum in Ω [Proposition (P.19)]: if $E_j \in \mathscr{E}_{\alpha_j}$ for $j \in \mathbf{N}$ and $\alpha = \sup\{\alpha_j\}$, then $E_j \in \mathscr{E}_\alpha$ for all j and hence $\bigcup_1^\infty E_j \in \mathscr{E}_\beta$ where β is the successor of α. □

Combining this result with Proposition (P.14), we see that if $\text{card}(\mathbf{N}) \leq \text{card}(\mathscr{E}) \leq \mathfrak{c}$, then $\text{card}(\mathscr{M}(\mathscr{E})) = \mathfrak{c}$. (Cf. Exercise 3.)

Section 3. Some authors prefer to take the domains of measures to be σ-rings rather than σ-algebras (cf. Exercise 1). The reason is that in dealing with "very large" spaces one can avoid certain pathologies by not attempting to measure "very large" sets. However, this point of view also has technical disadvantages, and it is no longer much in favor.

Section 4. Carathéodory's theorem appears in [16]. Theorem (1.14) is commonly attributed to Hahn and/or Carathéodory, but it is originally due to Fréchet [37]. The proof via Carathéodory's theorem was discovered independently by Hahn [39] and Kolmogorov [59].

Section 5. Lebesgue originally defined the outer measure $m^*(E)$ of a set $E \subset \mathbf{R}$ in terms of countable coverings by intervals, as we have done. He then defined a bounded set E to be measurable if $m^*(E) + m^*((a,b)\setminus E) = b - a$, where (a,b) is an interval containing E, and an unbounded set to be measurable if its intersection with any bounded interval is measurable. Carathéodory's characterization of measurability, which is technically easier to work with, of course came later. For the equivalence of two definitions, see Exercise 19.

One should convince oneself that the remarkably fussy proof of Proposition (1.15) is necessary by contemplating the complicated ways in which an h-interval can be decomposed into a disjoint union of h-subintervals. In any such decomposition the collection of right endpoints of the subintervals, when ordered from right to left on the line, is a well ordered set, but it can be order isomorphic to any initial segment of the set of countable ordinals.

Lebesgue measure can be extended to a translation-invariant measure on σ-algebras which properly include $\mathscr{L}$; see Kakutani and Oxtoby [56]. For $n = 1, 2$, Lebesgue measure on $\mathbf{R}^n$ can also be extended to a finitely additive measure on $\mathscr{P}(\mathbf{R}^n)$ which is invariant under translations and rotations; see Banach [5].

In connection with the existence of nonmeasurable sets, Solovay [99] has proved a remarkable theorem which says in effect that it is impossible to construct Lebesgue nonmeasurable sets without using the axiom of choice for uncountable collections of sets. (The precise statement of the theorem involves some technical points of axiomatic set theory, into which we shall not go here.) From the point of view of the working analyst, the effect of Solovay's theorem is to reaffirm the adequacy of the Lebesgue theory for all practical purposes.

The construction of the Cantor set admits an obvious generalization in which one successively removes various fractional parts of intervals other than the middle thirds. More precisely, let $\{\xi_n\}_0^\infty$ be a sequence of positive numbers such that $\xi_0 = 1$ and $\xi_n > 2\xi_{n+1}$ for all n. Remove from $[0,1]$ the open middle interval of length $\xi_0 - 2\xi_1$ (by an "open middle interval" of an interval I we mean an open interval having the same midpoint as I), obtaining a set B_1 which is the union of the two intervals $[0, \xi_1]$ and $[1 - \xi_1, 1]$. Proceeding inductively, having constructed B_n, remove from each of its 2^n constituent

intervals the open middle intervals of length $\xi_n - 2\xi_{n+1}$ to obtain a set B_{n+1} which is the union of 2^{n+1} closed intervals of length ξ_{n+1}. Finally, let $B = \bigcap_1^\infty B_n$; B is called a **generalized Cantor set.** Just like the ordinary Cantor set, B is compact, nowhere dense, and totally disconnected but with no isolated points, and $m(B) = \lim_{n\to\infty} 2^n\xi_n$. In particular, if we take $\xi_n = ((1 - \theta)/2)^n$ where $0 < \theta < 1$, so that at each stage we are removing the fixed fraction θ of each interval, then $m(B) = 0$. (The ordinary Cantor set is the case $\theta = 1/3$.) On the other hand, if we take $\xi_n = \alpha 2^{-n}$ where $0 < \alpha < 1$, then $m(B) = \alpha$. We thus obtain closed, nowhere dense subsets of $[0,1]$ whose measure is positive, indeed, whose measure is arbitrarily close to that of $[0,1]$. (There does not exist a nowhere dense set $E \subset [0,1]$ such that $m(E) = 1$, however, for $[0,1]\setminus\overline{E}$ is open and nonempty, and it therefore has positive measure.)

2

INTEGRATION

In the classical theory of integration on $\mathbf{R}$, $\int_a^b f(x)\,dx$ is defined as a limit of Riemann sums, which are integrals of functions which approximate f and are constant on subintervals of $[a, b]$. Similarly, on any measure space there is an obvious notion of integral for functions which are, in a suitable sense, locally constant, and it can be extended to an integral for more general functions. In this chapter we develop the theory of integration on abstract measure spaces, paying particular attention to the Lebesgue integral on $\mathbf{R}^n$.

1. MEASURABLE FUNCTIONS

We begin our study of integration theory with a discussion of measurable mappings, which are the morphisms in the category of measurable spaces.

We recall that any mapping $f\colon X \to Y$ between two sets induces a mapping $f^{-1}\colon \mathscr{P}(X) \to \mathscr{P}(Y)$ by $f^{-1}(E) = \{x \in X\colon f(x) \in E\}$ which commutes with unions, intersections, and complements. Thus if $\mathscr{N}$ is a σ-algebra on Y, $\{f^{-1}(E)\colon E \in \mathscr{N}\}$ is a σ-algebra on X. If $(X, \mathscr{M})$ and $(Y, \mathscr{N})$ are measurable spaces, a mapping $f\colon X \to Y$ is called **$(\mathscr{M}, \mathscr{N})$-measurable** (or just **measurable** if $\mathscr{M}$ and $\mathscr{N}$ are understood) if $f^{-1}(E) \in \mathscr{M}$ for all $E \in \mathscr{N}$. It is obvious that the composition of measurable mappings is measurable; that is, if $f\colon X \to Y$ is $(\mathscr{M}, \mathscr{N})$-measurable and $g\colon Y \to Z$ is $(\mathscr{N}, \mathscr{O})$-measurable, then $g \circ f$ is $(\mathscr{M}, \mathscr{O})$-measurable.

(2.1) Proposition. If $\mathscr{N}$ is generated by $\mathscr{E}$, then f is $(\mathscr{M}, \mathscr{N})$-measurable iff $f^{-1}(E) \in \mathscr{M}$ for all $E \in \mathscr{E}$.

Proof. The "only if" implication is trivial. For the converse, observe that $\{E \subset Y\colon f^{-1}(E) \in \mathscr{M}\}$ is a σ-algebra which contains $\mathscr{E}$; it therefore contains $\mathscr{N}$. □

(2.2) Corollary. If X and Y are metric (or topological) spaces, every continuous $f\colon X \to Y$ is $(\mathscr{B}_X, \mathscr{B}_Y)$-measurable.

Proof. f is continuous iff $f^{-1}(U)$ is open in X for every open $U \subset Y$, so the assertion follows from Proposition (2.1). □

If $(X, \mathscr{M})$ is a measurable space, a real- or complex-valued function f on X will be called **$\mathscr{M}$-measurable**, or just **measurable**, if it is $(\mathscr{M}, \mathscr{B}_\mathbf{R})$ or $(\mathscr{M}, \mathscr{B}_\mathbf{C})$ measurable. $\mathscr{B}_\mathbf{R}$ or $\mathscr{B}_\mathbf{C}$ is *always* understood as the σ-algebra on the range space unless otherwise specified. Special case: $f\colon \mathbf{R} \to \mathbf{R}$ is **Lebesgue** (resp. **Borel**) **measurable** if it is $(\mathscr{L}, \mathscr{B}_\mathbf{R})$ [resp. $(\mathscr{B}_\mathbf{R}, \mathscr{B}_\mathbf{R})$] measurable; likewise for $f\colon \mathbf{R} \to \mathbf{C}$.

Warning: If $f, g\colon \mathbf{R} \to \mathbf{R}$ are Lebesgue measurable, it does not follow that $f \circ g$ is Lebesgue measurable, even if g is assumed continuous. [If $E \in \mathscr{B}_\mathbf{R}$ we have $f^{-1}(E) \in \mathscr{L}$, but unless $f^{-1}(E) \in \mathscr{B}_\mathbf{R}$ there is no guarantee that $g^{-1}(f^{-1}(E))$ will be in $\mathscr{L}$. See Exercise 9.] However, if f is Borel measurable, then $f \circ g$ is Lebesgue or Borel measurable whenever g is.

(2.3) Proposition. $f\colon X \to \mathbf{R}$ is $\mathscr{M}$-measurable iff $f^{-1}((a, \infty)) \in \mathscr{M}$ $(a \in \mathbf{R})$ iff $f^{-1}([a, \infty)) \in \mathscr{M}$ $(a \in \mathbf{R})$ iff $f^{-1}((-\infty, a)) \in \mathscr{M}$ $(a \in \mathbf{R})$ iff $f^{-1}((-\infty, a]) \in \mathscr{M}$ $(a \in \mathbf{R})$.

Proof. This follows from Propositions (1.2) and (2.1). □

Sometimes we wish to consider measurability on subsets of X. If $(X, \mathscr{M})$ is a measurable space, f is a function on X, and $A \in \mathscr{M}$, we say that f is **measurable on** A if $f^{-1}(E) \cap A \in \mathscr{M}$ for all Borel sets E. (Equivalently, $f|A$ is $\mathscr{M}_A$-measurable where $\mathscr{M}_A = \{B \cap A\colon B \in \mathscr{M}\}$.)

If X is a set, $\{(Y_\alpha, \mathscr{N}_\alpha)\}_{\alpha \in A}$ is a family of measurable spaces, and $f_\alpha\colon X \to Y_\alpha$ is a map for each $\alpha \in A$, there is a unique smallest σ-algebra on X with respect to which the f_α's are all measurable, namely, the σ-algebra generated by the sets $f_\alpha^{-1}(E_\alpha)$ for $E_\alpha \in \mathscr{N}_\alpha$ and $\alpha \in A$. It is called the σ-algebra **generated** by $\{f_\alpha\}_{\alpha \in A}$. In particular, if $X = \prod_{\alpha \in A} Y_\alpha$, we see that the product σ-algebra on X, as defined in Section 1.2, is the σ-algebra generated by the projections $\pi_\alpha\colon X \to Y_\alpha$.

(2.4) Proposition. Let $(X, \mathscr{M})$ and $(Y_\alpha, \mathscr{N}_\alpha)$ $(\alpha \in A)$ be measurable spaces, $Y = \prod_{\alpha \in A} Y_\alpha$, $\mathscr{N} = \otimes_{\alpha \in A} \mathscr{N}_\alpha$, $\pi_\alpha\colon Y \to Y_\alpha$ the projections. Then $f\colon X \to Y$ is $(\mathscr{M}, \mathscr{N})$-measurable iff $f_\alpha = \pi_\alpha \circ f$ is $(\mathscr{M}, \mathscr{N}_\alpha)$-measurable for all α.

Proof. If f is measurable, so is f_α since the composition of measurable maps is measurable. Conversely, if each f_α is measurable, then for all $E_\alpha \in \mathscr{N}_\alpha$, $f^{-1}(\pi_\alpha^{-1}(E_\alpha)) = f_\alpha^{-1}(E_\alpha) \in \mathscr{M}$, whence f is measurable by Proposition (2.1). □

(2.5) Corollary. $f\colon X \to \mathbf{C}$ is $\mathscr{M}$-measurable iff $\operatorname{Re} f$ and $\operatorname{Im} f$ are $\mathscr{M}$-measurable.

Proof. This follows since $\mathscr{B}_\mathbf{C} = \mathscr{B}_{\mathbf{R}^2} = \mathscr{B}_\mathbf{R} \otimes \mathscr{B}_\mathbf{R}$ by Theorem (1.5). □

We now establish that measurability is preserved under the familiar algebraic and limiting operations.

(2.6) Proposition. If $f, g: X \to \mathbf{C}$ are $\mathcal{M}$-measurable, then so are $f + g$ and fg.

Proof. Define $F: X \to \mathbf{C} \times \mathbf{C}$, $\varphi: \mathbf{C} \times \mathbf{C} \to \mathbf{C}$, and $\psi: \mathbf{C} \times \mathbf{C} \to \mathbf{C}$ by $F = (f, g)$, $\varphi(x, y) = x + y$, $\psi(x, y) = xy$. Since $\mathcal{B}_{\mathbf{C}\times\mathbf{C}} = \mathcal{B}_{\mathbf{C}} \otimes \mathcal{B}_{\mathbf{C}}$ by Theorem (1.5), F is $(\mathcal{M}, \mathcal{B}_{\mathbf{C}\times\mathbf{C}})$-measurable by Proposition (2.4), whereas φ and ψ are $(\mathcal{B}_{\mathbf{C}\times\mathbf{C}}, \mathcal{B}_{\mathbf{C}})$-measurable by Corollary (2.2). Thus $f + g = \varphi \circ F$ and $fg = \psi \circ F$ are $\mathcal{M}$-measurable. □

It is sometimes convenient to consider functions with values in the extended real number system $\overline{\mathbf{R}} = [-\infty, \infty]$. We define Borel sets in $\overline{\mathbf{R}}$ by $\mathcal{B}_{\overline{\mathbf{R}}} = \{E \subset \overline{\mathbf{R}}: E \cap \mathbf{R} \in \mathcal{B}_{\mathbf{R}}\}$. [This coincides with the usual definition of the Borel σ-algebra if we make $\overline{\mathbf{R}}$ into a metric space with metric $\rho(x, y) = |A(x) - A(y)|$, where $A(x) = \arctan x$.] It is easily verified that $\mathcal{B}_{\overline{\mathbf{R}}}$ is generated by the rays $(a, \infty]$ or $[-\infty, a)$ $(a \in \mathbf{R})$, and we define $f: X \to \overline{\mathbf{R}}$ to be $\mathcal{M}$-measurable if it is $(\mathcal{M}, \mathcal{B}_{\overline{\mathbf{R}}})$-measurable. Proposition (2.6) remains valid for $\overline{\mathbf{R}}$-valued functions, provided one takes a little care with the indeterminate expressions $\infty - \infty$ and $0 \cdot \infty$. (Recall, however, that by convention we always define $0 \cdot \infty$ to be 0.) See Exercise 2.

(2.7) Proposition. If $\{f_j\}$ is a sequence of $\overline{\mathbf{R}}$-valued measurable functions on $(X, \mathcal{M})$, then the functions

$$g_1(x) = \sup_j f_j(x), \qquad g_2(x) = \inf_j f_j(x),$$

$$g_3(x) = \limsup_{j\to\infty} f_j(x), \qquad g_4(x) = \liminf_{j\to\infty} f_j(x)$$

are all measurable. If $f(x) = \lim_{j\to\infty} f_j(x)$ exists for every $x \in X$, then f is measurable.

Proof. We have $g_1^{-1}((a, \infty]) = \bigcup_1^\infty f_j^{-1}((a, \infty])$ and $g_2^{-1}([-\infty, a)) = \bigcup_1^\infty f_j^{-1}([-\infty, a))$, so g_1 and g_2 are measurable by Proposition (2.1). More generally, if $h_k(x) = \sup_{j>k} f_j(x)$ then h_k is measurable for each k, so $g_3 = \inf_{k\geq 1} h_k$ is measurable, and likewise for g_4. Finally, if f exists then $f = g_3 = g_4$, so f is measurable. □

(2.8) Corollary. If $f, g: X \to \overline{\mathbf{R}}$ are measurable, then so are $\max(f, g)$ and $\min(f, g)$.

(2.9) Corollary. If $\{f_j\}$ is a sequence of complex-valued measurable functions and $f(x) = \lim f_j(x)$ exists for all x, then f is measurable.

Proof. Apply Corollary (2.5). □

At this point we discuss two useful decompositions of functions. First, if $f: X \to \overline{\mathbf{R}}$, we define the **positive** and **negative parts** of f to be

$$f^+(x) = \max(f(x), 0), \qquad f^-(x) = \max(-f(x), 0).$$

Thus $f = f^+ - f^-$. If f is a measurable, so are f^+ and f^-, by Corollary (2.8). Second, if $f: X \to \mathbf{C}$, we have its **polar decomposition**

$$f = (\operatorname{sgn} f)|f|, \text{ where } \operatorname{sgn} z = z/|z| \text{ if } z \neq 0, \text{ and } \operatorname{sgn} 0 = 0.$$

Again, if f is measurable, so are $|f|$ and $\operatorname{sgn} f$. [$z \to |z|$ is continuous, and $z \to \operatorname{sgn} z$ is continuous except at the origin. If $U \subset \mathbf{C}$ is open, $\operatorname{sgn}^{-1}(U)$ is either open or equal to $V \cup \{0\}$ where V is open, so sgn is Borel measurable.]

Let $(X, \mathcal{M})$ be a measurable space. If $E \subset X$, the **characteristic function** χ_E of E (sometimes called the **indicator function** of E and denoted by 1_E) is defined by $\chi_E(x) = 1$ if $x \in E$, $\chi_E(x) = 0$ if $x \in E^c$. It is easily checked that χ_E is measurable iff $E \in \mathcal{M}$. A **simple function** on X is a finite linear combination, with complex coefficients, of characteristic functions of sets in $\mathcal{M}$. (We do not allow simple functions to assume the values $\pm\infty$.) Equivalently, $f: X \to \mathbf{C}$ is simple iff f is measurable and range(f) is a finite subset of $\mathbf{C}$. Indeed, if range$(f) = \{z_1, \ldots, z_n\}$, let $E_j = f^{-1}(\{z_j\})$: then $f = \sum_1^n z_j \chi_{E_j}$. We call this the **standard representation** for f. It exhibits f as a linear combination, with distinct coefficients, of disjoint sets whose union is X.

It is clear that if f and g are simple functions, then so are $f + g$ and fg. We now show that arbitrary measurable functions can be approximated in a nice way by simple functions. (The reader should draw a picture of the constructions in this theorem.)

(2.10) Theorem. Let $(X, \mathcal{M})$ be a measurable space.

(a) If $f: X \to [0, \infty]$ is measurable, there is a sequence $\{\varphi_n\}$ of simple functions such that $0 \leq \varphi_1 \leq \varphi_2 \leq \cdots \leq f$, $\varphi_n \to f$ pointwise, and $\varphi_n \to f$ uniformly on any set on which f is bounded.

(b) If $f: X \to \mathbf{C}$ is measurable, there is a sequence $\{\varphi_n\}$ of simple functions such that $0 \leq |\varphi_1| \leq |\varphi_2| \leq \cdots \leq |f|$, $\varphi_n \to f$ pointwise, and $\varphi_n \to f$ uniformly on any set on which f is bounded.

Proof. (a) If $n \in \mathbf{N}$ and $0 \leq k \leq 2^{2n} - 1$, let

$$E_n^k = f^{-1}((k2^{-n}, (k+1)2^{-n}]) \quad \text{and} \quad F_n = f^{-1}((2^n, \infty]),$$

and define

$$\varphi_n = \sum_{k=0}^{2^{2n}-1} k2^{-n}\chi_{E_n^k} + 2^n\chi_{F_n}.$$

It is easily checked that $\varphi_n \le \varphi_{n+1}$ for all n, and $0 \le f - \varphi_n \le 2^{-n}$ on the set where $f \le 2^n$. The result therefore follows.

(b) If $f = g + ih$, we can apply part (a) to the positive and negative parts of g and h, obtaining sequences ψ_n^+, ψ_n^-, ζ_n^+, ζ_n^- of nonnegative simple functions which increase to g^+, g^-, h^+, h^-. Let $\varphi_n = \psi_n^+ - \psi_n^- + i(\zeta_n^+ - \zeta_n^-)$; it is then a simple exercise to verify that φ_n has the desired properties. □

If μ is a measure on $(X, \mathcal{M})$, we may wish to except μ-null sets from consideration in studying measurable functions. In this respect life is a bit simpler if μ is complete.

(2.11) Proposition. If μ is complete, then

(a) if f is measurable and $f = g$ a.e., then g is measurable;

(b) if f_n is measurable for $n \in \mathbf{N}$ and $f_n \to f$ a.e., then f is measurable.

If μ is not complete, (a) and (b) are in general false.

The proof is left to the reader (Exercise 16).

However, the following result shows that one is unlikely to commit any serious blunders by forgetting to worry about completeness of the measure.

(2.12) Proposition. Let $(X, \mathcal{M}, \mu)$ be a measure space and let $(X, \bar{\mathcal{M}}, \bar{\mu})$ be its completion. If f is an $\bar{\mathcal{M}}$-measurable function on X, there is an $\mathcal{M}$-measurable function g such that $f = g$ $\bar{\mu}$-almost everywhere.

Proof. This is obvious if $f = \chi_E$ where $E \in \bar{\mathcal{M}}$, and hence if f is an $\bar{\mathcal{M}}$-measurable simple function. For the general case, choose a sequence $\{\varphi_n\}$ of $\bar{\mathcal{M}}$-measurable simple functions which converge pointwise to f according to Theorem (2.10), and for each n let ψ_n be an $\mathcal{M}$-measurable simple function with $\psi_n = \varphi_n$ except on a set $E_n \in \bar{\mathcal{M}}$ with $\bar{\mu}(E_n) = 0$. Choose $N \in \mathcal{M}$ such that $\mu(N) = 0$ and $N \supset \bigcup_1^\infty E_n$, and set $g = \lim \psi_n \chi_{N^c}$. Then g is $\mathcal{M}$-measurable by Corollary (2.9), and $g = f$ on N^c. □

EXERCISES

In Exercises 1–7, $(X, \mathcal{M})$ is a measurable space.

1. Let $f: X \to \bar{\mathbf{R}}$ and $Y = f^{-1}(\mathbf{R})$. Then f is measurable iff $f^{-1}(\{-\infty\}) \in \mathcal{M}$, $f^{-1}(\{\infty\}) \in \mathcal{M}$, and f is measurable on Y.
2. Suppose, $f, g: X \to \bar{\mathbf{R}}$ are measurable.
   (a) fg is measurable [where $0 \cdot (\pm\infty) = 0$].
   (b) Fix $\alpha \in \bar{\mathbf{R}}$ and define $h(x) = \alpha$ if $f(x) = -g(x) = \pm\infty$ and $h(x) = f(x) + g(x)$ otherwise. Then h is measurable.

3. If $\{f_n\}$ is a sequence of measurable functions on X, then $\{x: \lim f_n(x) \text{ exists}\}$ is in $\mathcal{M}$.
4. If $f: X \to \bar{\mathbf{R}}$ and $f^{-1}((r, \infty]) \in \mathcal{M}$ for each $r \in \mathbf{Q}$, then f is measurable.
5. If $X = A \cup B$ where $A, B \in \mathcal{M}$, a function f on X is measurable iff f is measurable on A and on B.
6. The supremum of an uncountable family of measurable $\bar{\mathbf{R}}$-valued functions on X can fail to be measurable.
7. Suppose that for each $\alpha \in \mathbf{R}$ we are given a set $E_\alpha \in \mathcal{M}$ such that $E_\alpha \subset E_\beta$ whenever $\alpha < \beta$, $\bigcup_{\alpha \in \mathbf{R}} E_\alpha = X$, and $\bigcap_{\alpha \in \mathbf{R}} E_\alpha = \varnothing$. Then there is a measurable function $f: X \to \mathbf{R}$ such that $f(x) \le \alpha$ on E_α and $f(x) \ge \alpha$ on E_α^c for every α. (Use Exercise 4.)
8. If $f: \mathbf{R} \to \mathbf{R}$ is monotone, then f is Borel measurable.
9. Let $f: [0,1] \to [0,1]$ be the Cantor function (Section 1.5), and let $g: [0,1] \to [0,2]$ be defined by $g(x) = f(x) + x$.
   (a) g maps $[0,1]$ bijectively onto $[0,2]$, and $h = g^{-1}$ is continuous from $[0,2]$ to $[0,1]$.
   (b) If C is the Cantor set, $m(g(C)) = 1$.
   (c) By Exercise 1.29, $g(C)$ contains a Lebesgue nonmeasurable set A. Let $B = g^{-1}(A)$. Then B is Lebesgue measurable but not Borel.
   (d) There exist a Lebesgue measurable function F and a continuous function G on $\mathbf{R}$ such that $F \circ G$ is not Lebesgue measurable.
10. Prove Proposition (2.11).
11. Suppose f is a function on $\mathbf{R} \times \mathbf{R}^k$ such that $f(x, \cdot)$ is Borel measurable for each $x \in \mathbf{R}$ and $f(\cdot, y)$ is continuous for each $y \in \mathbf{R}^k$. For $n \in \mathbf{N}$, define f_n as follows. Let $a_i = i/n$ $(i \in \mathbf{Z})$, and for $a_i \le x \le a_{i+1}$ let

$$f_n(x, y) = \frac{f(a_{i+1}, y)(x - a_i) - f(a_i, y)(x - a_{i+1})}{a_{i+1} - a_i}.$$

Then f_n is Borel measurable on $\mathbf{R} \times \mathbf{R}^k$ and $f_n \to f$ pointwise; hence f is Borel measurable on $\mathbf{R} \times \mathbf{R}^k$. Conclude by induction that every function on $\mathbf{R}^n$ which is continuous in each variable separately is Borel measurable.

2. INTEGRATION OF NONNEGATIVE FUNCTIONS

In this section we fix a measure space $(X, \mathcal{M}, \mu)$ and we denote by L^+ the space of all measurable functions from X to $[0, \infty]$. If φ is a simple function in L^+ with standard representation $\varphi = \sum_1^n a_j \chi_{E_j}$, we define the **integral** of φ with respect to μ to be

$$\int \varphi \, d\mu = \sum_1^n a_j \mu(E_j)$$

(with the convention, as always, that $0 \cdot \infty = 0$). We note that $\int \varphi \, d\mu$ may equal ∞. When there is no danger of confusion, we shall also write $\int \varphi$ instead of $\int \varphi \, d\mu$. Also, it is sometimes convenient to display the argument of φ explicitly, especially when $\varphi(x)$ is given by a formula in terms of x or when there are several variables involved; in this case we shall use the notation $\int \varphi(x)\, d\mu(x)$ for $\int \varphi \, d\mu$. [Some authors prefer to write $\int \varphi(x) \mu(dx)$.] Finally, if $A \in \mathcal{M}$, then $\varphi \chi_A$ is also simple (viz., $\varphi \chi_A = \sum a_j \chi_{A \cap E_j}$), and we define $\int_A \varphi \, d\mu$ [or $\int_A \varphi$, or $\int_A \varphi(x)\, d\mu(x)$] to be $\int \varphi \chi_A \, d\mu$. The same notational conventions will apply also to the more general integrals we define below.

(2.13) Proposition. Let φ and ψ be simple functions in L^+.

(a) If $c \geq 0$, $\int c\varphi = c \int \varphi$.

(b) $\int(\varphi + \psi) = \int \varphi + \int \psi$.

(c) If $\varphi \leq \psi$, then $\int \varphi \leq \int \psi$.

(d) The map $A \to \int_A \varphi$ is a measure on $\mathcal{M}$.

Proof. (a) is trivial. For (b), let $\sum_1^n a_j \chi_{E_j}$ and $\sum_1^m b_k \chi_{F_k}$ be the standard representations of φ and ψ. Since $E_j = \bigcup_{k=1}^m (E_j \cap F_k)$ and $F_k = \bigcup_{j=1}^n (E_j \cap F_k)$, where the unions are disjoint, finite additivity of μ implies that

$$\int \varphi + \int \psi = \sum_{j,k} (a_j + b_k) \mu(E_j \cap F_k),$$

and the same reasoning shows that the sum on the right equals $\int(\varphi + \psi)$. Moreover, if $\varphi \leq \psi$ then $a_j \leq b_k$ whenever $E_j \cap F_k \neq \varnothing$, so

$$\int \varphi = \sum_{j,k} a_j \mu(E_j \cap F_k) \leq \sum_{j,k} b_k \mu(E_j \cap F_k) = \int \psi,$$

which proves (c). Finally, if $\{A_k\}_1^\infty$ is a disjoint sequence and $A = \bigcup_1^\infty A_k$,

$$\int_A \varphi = \sum_j a_j \mu(A \cap E_j) = \sum_{j,k} a_j \mu(A_k \cap E_j) = \sum_k \int_{A_k} \varphi,$$

which establishes (d). □

We now extend the integral to all functions $f \in L^+$ by defining

$$\int f \, d\mu = \sup\left\{ \int \varphi \, d\mu : 0 \leq \varphi \leq f, \varphi \text{ simple} \right\}.$$

By Proposition (2.13c), the two definitions of $\int f$ agree when f is simple. Moreover, it is obvious from the definition that $\int f \leq \int g$ whenever $f \leq g$, and that $\int cf = c \int f$ for all $c \geq 0$. The next step is to establish one of the fundamental convergence theorems, which, in conjunction with Theorem (2.10), shows that to compute $\int f$ it suffices to examine a sequence of simple functions which increase to f.

(2.14) The Monotone Convergence Theorem. If $\{f_n\}$ is a sequence in L^+ such that $f_j \leq f_{j+1}$ for all j, and $f = \lim_{n \to \infty} f_n$ $(= \sup_n f_n)$, then $\int f = \lim_{n \to \infty} \int f_n$.

Proof. $\{\int f_n\}$ is an increasing sequence of numbers, so its limit exists (possibly equal to ∞). Moreover, $\int f_n \leq \int f$ for all n, so $\lim \int f_n \leq \int f$. To establish the reverse inequality, fix $\alpha \in (0,1)$, let φ be a simple function with $0 \leq \varphi \leq f$, and let $E_n = \{x: f_n(x) \geq \alpha \varphi(x)\}$. Then $\{E_n\}$ is an increasing sequence of measurable sets whose union is X, and we have $\int f_n \geq \int_{E_n} f_n \geq \alpha \int_{E_n} \varphi$. By Proposition (2.13d) and Theorem (1.8c), $\lim \int_{E_n} \varphi = \int \varphi$, and hence $\lim \int f_n \geq \alpha \int \varphi$. Since this is true for all $\alpha < 1$, it remains true for $\alpha = 1$, and taking the supremum over all simple $\varphi \leq f$ we obtain $\lim \int f_n \geq \int f$. □

With this result in hand, we can establish the additivity of the integral.

(2.15) Theorem. If $\{f_n\}$ is a finite or infinite sequence in L^+ and $f = \sum_n f_n$, then $\int f = \sum_n \int f_n$.

Proof. First consider two functions f_1 and f_2. By Theorem (2.10) we can find sequences $\{\varphi_j\}$ and $\{\psi_j\}$ of nonnegative simple functions which increase to f_1 and f_2. Then $\{\varphi_j + \psi_j\}$ increases to $f_1 + f_2$, so by the monotone convergence theorem,

$$\int (f_1 + f_2) = \lim \int (\varphi_j + \psi_j) = \lim \int \varphi_j + \lim \int \psi_j = \int f_1 + \int f_2.$$

Now, by induction, $\int \sum_1^N f_n = \sum_1^N \int f_n$, so letting $N \to \infty$ and applying the monotone convergence theorem again, we obtain $\int \sum_1^\infty f_n = \sum_1^\infty \int f_n$. □

(2.16) Proposition. If $f \in L^+$, then $\int f = 0$ iff $f = 0$ a.e.

Proof. This is obvious if f is simple: if $f = \sum_1^n a_j \chi_{E_j}$, $a_j \geq 0$, then $\int f = 0$ iff for each j either $a_j = 0$ or $\mu(E_j) = 0$. In general, if $f = 0$ a.e. and φ is simple with $0 \leq \varphi \leq f$, then $\varphi = 0$ a.e.; hence $\int f = \sup_{\varphi \leq f} \int \varphi = 0$. On the other hand, $\{x: f(x) > 0\} = \bigcup_1^\infty E_n$ where $E_n = \{x: f(x) > n^{-1}\}$, so if it is false that $f = 0$ a.e., we must have $\mu(E_n) > 0$ for some n. But then $f > n^{-1} \chi_{E_n}$, so $\int f \geq n^{-1} \mu(E_n) > 0$. □

(2.17) Corollary. If $\{f_n\} \subset L^+$, $f \in L^+$, and $f_n(x)$ increases to $f(x)$ for a.e. x, then $\int f = \lim \int f_n$.

Proof. If $f_n(x)$ increases to $f(x)$ for $x \in E$ where $\mu(E^c) = 0$, then $f = f\chi_E$ and $f_n = f_n \chi_E$ a.e., so $\int f = \int f \chi_E = \lim \int f_n \chi_E = \lim \int f_n$. □

The hypothesis that the sequence $\{f_n\}$ be increasing, at least a.e., is essential for the monotone convergence theorem. For example, if X is $\mathbf{R}$ and μ is Lebesgue measure we have $\chi_{(n,n+1)} \to 0$ pointwise but $\int \chi_{(n,n+1)} = 1$ for all n. However, the integral of the limit can never exceed the limit of the integrals.

(2.18) Fatou's Lemma. If $\{f_n\}$ is any sequence in L^+, then

$$\int (\liminf f_n) \leq \liminf \int f_n.$$

Proof. For each k we have $\inf_{n \geq k} f_n \leq f_j$ for $j \geq k$, hence $\int \inf_{n \geq k} f_n \leq \int f_j$ for $j \geq k$, hence $\int \inf_{n \geq k} f_n \leq \inf_{j \geq k} \int f_j$. Now let $k \to \infty$ and apply the monotone convergence theorem:

$$\int (\liminf f_n) = \lim_{k \to \infty} \int (\inf_{n \geq k} f_n) \leq \liminf \int f_n. \qquad \square$$

(2.19) Corollary. If $\{f_n\} \subset L^+$, $f \in L^+$, and $f_n \to f$ a.e., then $\int f \leq \liminf \int f_n$.

Proof. If $f_n \to f$ everywhere the result is immediate from Fatou's lemma, and this can be achieved by modifying f_n and f on a null set without affecting the integrals, by Proposition (2.16). $\square$

(2.20) Proposition. If $f \in L^+$ and $\int f < \infty$, then $\{x\colon f(x) = \infty\}$ is a null set and $\{x\colon f(x) > 0\}$ is σ-finite.

The proof is left to the reader (Exercise 12).

EXERCISES

12. Prove Proposition (2.20). [Cf. Proposition (P.20), where a special case is proved.]

13. Suppose $\{f_n\} \subset L^+$, $f_n \to f$ pointwise, and $\int f = \lim \int f_n < \infty$. Then $\int_E f = \lim \int_E f_n$ for all $E \in \mathcal{M}$. However, this need not be true if $\int f = \lim \int f_n = \infty$.

14. If $f \in L^+$, let $\lambda(E) = \int_E f\,d\mu$ for $E \in \mathcal{M}$. Then λ is a measure on $\mathcal{M}$, and for any $g \in L^+$, $\int g\,d\lambda = \int fg\,d\mu$. (First suppose g is simple.)

15. If $\{f_n\} \subset L^+$, f_n decreases pointwise to f, and $\int f_k < \infty$ for some k, then $\int f = \lim \int f_n$.

16. If $f \in L^+$ and $\int f < \infty$, for every $\varepsilon > 0$ there exists $E \in \mathcal{M}$ with $\mu(E) < \infty$ and $\int_E f > \int f - \varepsilon$.

17. It is possible to prove Fatou's lemma without using the monotone convergence theorem. Show that the monotone convergence theorem can then be proved as a corollary of Fatou's lemma.

18. Find a sequence $\{f_n\}$ of Borel measurable functions on $\mathbf{R}$ which decreases uniformly to zero on $\mathbf{R}$, but $\int f_n\,dm = \infty$ for all n. Also, find a sequence $\{g_n\}$ of Borel measurable functions on $[0,1]$ such that $g_n \to 0$ pointwise but $\int g_n\,dm = 1$ for all n.

3. INTEGRATION OF COMPLEX FUNCTIONS

We continue to work on a fixed measure space $(X, \mathcal{M}, \mu)$. The integral defined in the previous section can be extended to real-valued measurable functions f in an obvious way: if f^+ and f^- are the positive and negative parts of f and at least one of $\int f^+$ and $\int f^-$ is finite, we define

$$\int f = \int f^+ - \int f^-.$$

We shall be mainly concerned with the case where $\int f^+$ and $\int f^-$ are both finite: we then say that f is **integrable**. Since $f^{\pm} \leq |f| \leq f^+ + f^-$, it is clear that f is integrable iff $\int |f| < \infty$.

(2.21) Proposition. The set of integrable real-valued functions is a real vector space, and the integral is a linear functional on it.

Proof. The first assertion follows from the fact that $|af + bg| \leq |a||f| + |b||g|$, and it is easy to check that $\int af = a \int f$ for any $a \in \mathbf{R}$. If f, g are integrable, let $h = f + g$. Then $h^+ - h^- = f^+ - f^- + g^+ - g^-$, so $h^+ + f^- + g^- = h^- + f^+ + g^+$. By Theorem (2.15),

$$\int h^+ + \int f^- + \int g^- = \int h^- + \int f^+ + \int g^+,$$

so that $\int h = \int f + \int g$. $\square$

Next, if f is a complex-valued measurable function, we say that f is **integrable** if $\int |f| < \infty$. (More generally, if $E \in \mathcal{M}$, f is **integrable on** E if $\int_E |f| < \infty$.) Since $|f| \leq |\mathrm{Re}\, f| + |\mathrm{Im}\, f| \leq 2|f|$, f is integrable iff $\mathrm{Re}\, f$ and $\mathrm{Im}\, f$ are integrable, and in this case we define

$$\int f = \int \mathrm{Re}\, f + i \int \mathrm{Im}\, f.$$

It follows easily that the space of complex-valued integrable functions is a complex vector space and that the integral is a complex linear functional on it.

We denote this space—provisionally—by $L^1(\mu)$ [or $L^1(X, \mu)$, or $L^1(X)$, or simply L^1, depending on the context]. The superscript 1 is standard notation, but it will not assume any significance for us until Chapter 6.

(2.22) Proposition. If $f \in L^1$, then $|\int f| \le \int |f|$.

Proof. This is trivial if $\int f = 0$ and almost trivial if f is real, since

$$\left|\int f\right| = \left|\int f^+ - \int f^-\right| \le \int f^+ + \int f^- = \int |f|.$$

If f is complex and $\int f \neq 0$, let $\alpha = \overline{\operatorname{sgn} \int f}$. Then

$$\left|\int f\right| = \alpha \int f = \int \alpha f = \operatorname{Re} \int \alpha f = \int \operatorname{Re}(\alpha f) \le \int |\operatorname{Re}(\alpha f)| \le \int |\alpha f| = \int |f|.$$

□

(2.23) Proposition.

(a) If $f \in L^1$, then $\{x: f(x) \neq 0\}$ is σ-finite.

(b) If $f, g \in L^1$, then $\int_E f = \int_E g$ for all $E \in \mathcal{M}$ iff $\int |f - g| = 0$ iff $f = g$ a.e.

Proof. (a) and the second equivalence in (b) follow from Proposition (2.20). If $\int |f - g| = 0$, then by Proposition (2.22), for any $E \in \mathcal{M}$,

$$\left|\int_E f - \int_E g\right| \le \int \chi_E |f - g| \le \int |f - g| = 0.$$

On the other hand, if $u = \operatorname{Re}(f - g)$, $v = \operatorname{Im}(f - g)$, and it is false that $f = g$ a.e., then at least one of u^+, u^-, v^+, v^- must be nonzero on a set of positive measure. If, say, $E = \{x: u^+(x) > 0\}$ has positive measure, then $\operatorname{Re}(\int_E f - \int_E g) = \int_E u^+ > 0$ since $u^- = 0$ on E; likewise in the other cases. □

This proposition shows that for the purposes of integration it makes no difference if we alter functions on null sets. Indeed, one can integrate functions f which are only *defined* on a measurable set E whose complement is null simply by defining f to be zero (or anything else) on E^c. In this fashion we can treat $\bar{\mathbf{R}}$-valued functions which are finite a.e. as real-valued functions for the purposes of integration.

With this in mind, we shall find it more convenient to redefine $L^1(\mu)$ to be the set of equivalence classes of a.e.-defined integrable functions on X, where f is equivalent to g iff $f = g$ a.e. This new $L^1(\mu)$ is still a complex vector space (under pointwise a.e. addition and scalar multiplication), and it is also a metric space with metric $\rho(f, g) = \int |f - g|$. Moreover, under the new definition, if $\bar{\mu}$ is the completion of μ then $L^1(\bar{\mu})$ can be canonically identified with $L^1(\mu)$, by Proposition (2.12). Henceforth, we shall view $L^1(\mu)$ as a space of equivalence classes, but we shall still employ the notation "$f \in L^1(\mu)$" to mean that f is an a.e.-defined integrable function. This minor abuse of notation is universally accepted and almost never causes any confusion.

We now present the last of the three basic convergence theorems (the other two being the monotone convergence theorem and Fatou's lemma) and derive some useful consequences from it.

(2.24) The Dominated Convergence Theorem. Let $\{f_n\}$ be a sequence in L^1 such that (a) $f_n \to f$ a.e., and (b) there exists a nonnegative $g \in L^1$ such that $|f_n| \le g$ for all n. Then $f \in L^1$ and $\int f = \lim_{n\to\infty} \int f_n$.

Proof. f is measurable (perhaps after redefinition on a null set) by Propositions (2.11) and (2.12), and since $|f| \le g$ a.e. we have $f \in L^1$. By taking real and imaginary parts it suffices to assume that f_n and f are real-valued, in which case we have $g + f_n \ge 0$ a.e. and $g - f_n \ge 0$ a.e. Thus by Fatou's lemma,

$$\int g + \int f \le \liminf \int (g + f_n) = \int g + \liminf \int f_n,$$

$$\int g - \int f \le \liminf \int (g - f_n) = \int g - \limsup \int f_n.$$

Therefore, $\liminf \int f_n \ge \int f \ge \limsup \int f_n$, and the result follows. □

(2.25) Theorem. Suppose $\{f_j\}$ is a sequence in L^1 such that $\sum_1^\infty \int |f_j| < \infty$. Then $\sum_1^\infty f_j$ converges a.e. to an L^1 function, and $\int \sum_1^\infty f_j = \sum_1^\infty \int f_j$.

Proof. By Theorem (2.15), $\int \sum_1^\infty |f_j| = \sum_1^\infty \int |f_j| < \infty$, so $\sum_1^\infty |f_j| \in L^1$. In particular, by Proposition (2.20), $\sum_1^\infty |f_j(x)|$ is finite for a.e. x, and for each such x the series $\sum_1^\infty f_j(x)$ converges. Moreover, $|\sum_1^n f_j| \le \sum_1^\infty |f_j|$ for all n, so we can apply the dominated convergence theorem to the sequence of partial sums to obtain $\int \sum_1^\infty f_j = \sum_1^\infty \int f_j$. □

(2.26) Theorem. If $f \in L^1(\mu)$ and $\varepsilon > 0$, there is an integrable simple function $\varphi = \sum a_j \chi_{E_j}$ such that $\int |f - \varphi| \, d\mu < \varepsilon$. [Thus, integrable simple functions are dense in $L^1(\mu)$ in the L^1 metric.] If μ is a Lebesgue–Stieltjes measure on $\mathbf{R}$, the sets E_j in the definition of φ can be taken to be finite unions of open intervals; moreover, there is a continuous function g which vanishes outside a bounded interval such that $\int |f - g| \, d\mu < \varepsilon$.

Proof. Let $\{\varphi_n\}$ be as in Theorem (2.10b); then $\int |\varphi_n - f| < \varepsilon$ for n sufficiently large, by the dominated convergence theorem. If $\varphi_n = \sum a_j \chi_{E_j}$ where the E_j's are disjoint and the a_j's are nonzero, we observe that $\mu(E_j) = |a_j|^{-1} \int_{E_j} |\varphi| \le |a_j|^{-1} \int |f| < \infty$. Moreover, $\mu(E \Delta F) = \int |\chi_E - \chi_F|$, so if μ is a

Lebesgue–Stieltjes measure on $\mathbf{R}$, by Proposition (1.20) we can approximate χ_{E_j} arbitrarily closely in the L^1 metric by finite sums of functions χ_{I_k} where the I_k's are open intervals. Finally, if $I_k = (a, b)$ we can approximate χ_{I_k} in the L^1 metric by continuous functions which vanish outside (a, b). For example, let g be the function such that $g(x) = 0$ for $x \notin (a, b)$, $g(x) = 1$ for $x \in [a + n^{-1}, b - n^{-1}]$, and g is linear on $[a, a + n^{-1}]$ and on $[b - n^{-1}, b]$; then

$$\left|\int (g - \chi_{(a,b)})\right| \leq \mu((a, a + n^{-1})) + \mu((b - n^{-1}, b)) \to 0 \text{ as } n \to \infty.$$

□

(2.27) Theorem. Suppose $f: X \times [a, b] \to \mathbf{C}$ $(-\infty < a < b < \infty)$ and $f(\cdot, t)$: $X \to \mathbf{C}$ is integrable for each $t \in [a, b]$. Let $F(t) = \int_X f(x, t)\, d\mu(x)$.

(a) Suppose there exists $g \in L^1(\mu)$ such that $|f(x, t)| \leq g(x)$ for all x, t. If $\lim_{t \to t_0} f(x, t) = f(x, t_0)$ for every x, then $\lim_{t \to t_0} F(t) = F(t_0)$; in particular, if $f(x, \cdot)$ is continuous for each x, then F is continuous.

(b) Suppose $\partial f/\partial t$ exists and there is a $g \in L^1(\mu)$ such that $|(\partial f/\partial t)(x, t)| \leq g(x)$ for all x, t. Then F is differentiable and $F'(t) = \int (\partial f/\partial t)(x, t)\, d\mu(x)$.

Proof. For (a), apply the dominated convergence theorem to $f_n(x) = f(x, t_n)$ where $\{t_n\}$ is any sequence converging to t_0. For (b), observe that

$$\frac{\partial f}{\partial t}(x, t_0) = \lim h_n(x) \text{ where } h_n(x) = \frac{f(x, t_n) - f(x, t_0)}{t_n - t_0},$$

$\{t_n\}$ again being any sequence converging to t_0. It follows that $\partial f/\partial t$ is measurable, and by the mean value theorem,

$$|h_n(x)| \leq \sup_{t \in [a,b]} \left|\frac{\partial f}{\partial t}(x, t)\right| \leq g(x),$$

so the dominated convergence theorem can be invoked again to give

$$F'(t_0) = \lim \frac{F(t_n) - F(t_0)}{t_n - t_0} = \lim \int h_n(x)\, d\mu(x) = \int \frac{\partial f}{\partial t}(x, t)\, d\mu(x).$$

□

The device of using sequences converging to t_0 in the preceding proof is technically necessary because the dominated convergence theorem deals only with sequences of functions. However, in such situations we shall usually just say "let $t \to t_0$" with the understanding that sequential convergence is underlying the argument.

Theorem (2.27) can easily be extended to the case where the parameter t varies over a noncompact interval, since its conclusions are local in nature. For example, if $I \subset \mathbf{R}$ is any interval and f: $X \times I \to \mathbf{C}$, to conclude that F is differentiable on I and that $F'(t) = \int (\partial f/\partial t)(x, t)\, d\mu(x)$ it suffices to show that the hypotheses of part (b) are satisfied on any compact subinterval of I.

We conclude this section with a comparison of the Lebesgue and Riemann integrals on $\mathbf{R}$. We shall use Darboux's characterization of the Riemann integral in terms of upper and lower sums, which we now recall.

Let $[a, b]$ be a compact interval. By a **partition** of $[a, b]$ we shall mean a finite sequence $P = \{t_j\}_0^n$ such that $a = t_0 < t_1 < \cdots < t_n = b$. Let f be an arbitrary bounded real-valued function on $[a, b]$. For each partition P we define

$$S_P f = \sum_1^n M_j(t_j - t_{j-1}), \qquad s_P f = \sum_1^n m_j(t_j - t_{j-1}),$$

where M_j and m_j are the supremum and infimum of f on $(t_{j-1}, t_j]$. Then we define

$$\overline{\int_a^b} f(x)\, dx = \inf S_P f, \quad \underline{\int_a^b} f(x)\, dx = \sup s_P f,$$

the infimum and supremum being taken over all partitions P. If $\overline{\int_a^b} f(x)\, dx = \underline{\int_a^b} f(x)\, dx$, their common value is the **Riemann integral** $\int_a^b f(x)\, dx$, and f is called **Riemann integrable.**

(2.28) Theorem. Let f be a bounded real-valued function on $[a, b]$.

(a) If f is Riemann integrable, then f is Lebesgue measurable (and hence integrable on $[a, b]$ since it is bounded), and $\int_a^b f(x)\, dx = \int_{[a,b]} f\, dm$.

(b) f is Riemann integrable iff $\{x \in [a, b]$: f is discontinuous at $x\}$ has Lebesgue measure zero.

Proof. Suppose f is Riemann integrable. For each partition P let

$$G_P = \sum_1^n M_j \chi_{(t_{j-1}, t_j]}, \qquad g_P = \sum_1^n m_j \chi_{(t_{j-1}, t_j]}$$

(with the same notation as above), so that $S_P f = \int G_P\, dm$ and $s_P f = \int g_P\, dm$. There is a sequence $\{P_k\}$ of partitions whose mesh [i.e., $\max_j(t_j - t_{j-1})$] tends to zero, each of which is a refinement of the preceding one (so that g_{P_k} increases with k while G_{P_k} decreases), such that $S_{P_k} f$ and $s_{P_k} f$ converge to $\int_a^b f(x)\, dx$. Let $G = \lim G_{P_k}$, $g = \lim g_{P_k}$. Then $g \leq f \leq G$, and by the dominated convergence theorem, $\int G\, dm = \int g\, dm = \int_a^b f(x)\, dx$. Hence $\int (G - g)\, dm = 0$, so $G = g$ a.e. by Proposition (2.16), and thus $G = f$ a.e. Since G is measurable (being the limit of a sequence of simple functions) and

m is complete, f is measurable and $\int_{[a,b]} f\,dm = \int G\,dm = \int_a^b f(x)\,dx$. This proves (a), and the proof of (b) is outlined in Exercise 23. □

The (proper) Riemann integral is thus subsumed in the Lebesgue integral. Some improper Riemann integrals can be interpreted directly as Lebesgue integrals, but others still require a limiting procedure. For example, if f is Riemann integrable on $[0, b]$ for all $b > 0$ and Lebesgue integrable on $[0, \infty)$, then $\int_{[0,\infty)} f\,dm = \lim_{b\to\infty} \int_0^b f(x)\,dx$ (by the dominated convergence theorem), but the limit on the right can exist even when f is not integrable [for example, consider $f = \sum_1^\infty n^{-1}(-1)^n \chi_{(n,n+1)}$]. Henceforth we shall generally use the notation $\int f(x)\,dx$ for Lebesgue integrals.

A few remarks comparing the construction of the Riemann and Lebesgue integrals may be helpful. Let f be a bounded measurable function on $[a, b]$, and for simplicity let us assume that $f \geq 0$. To compute the Riemann integral of f one partitions the domain $[a, b]$ into subintervals and approximates f from above and below by functions which are constant on each subinterval. To compute the Lebesgue integral of f one picks a sequence of simple functions which increase to f. In particular, if one picks the sequence constructed in the proof of Theorem (2.10a), one is in effect partitioning the *range* of f into subintervals I_j and approximating f by a constant on each of the sets $f^{-1}(I_j)$. This procedure requires a more sophisticated theory of measure to begin with since the sets $f^{-1}(I_j)$ can be complicated, even when f is continuous; but it is better adapted to the particular f under consideration and therefore more flexible—and more susceptible to generalization. (In the Lebesgue theory, the assumption that f is measurable removes the necessity of considering both upper and lower approximations; however, the latter point of view can also be made to work in the abstract setting. See Exercise 24.)

The Lebesgue theory offers two real advantages over the Riemann theory. First, much more powerful convergence theorems, such as the monotone and dominated convergence theorems, are available. These not only yield results previously unobtainable but also reduce the labor in proving classical theorems. Second, a wider class of functions can be integrated. For example, if R is the set of rational numbers in $[0, 1]$, χ_R is not Riemann integrable, being everywhere discontinuous on $[0, 1]$, but it is Lebesgue integrable, and $\int \chi_R\,dm = 0$. (Actually, this is in some sense a trivial example since χ_R agrees a.e. with the constant function 0. For a more interesting example, see Exercise 25.) Of course, virtually all functions that one meets in classical analysis are Riemann integrable, so this added generality is rarely used in computing specific integrals. However, it has the crucial effect that various metric spaces of functions whose metrics are defined in terms of integrals are *complete* when Lebesgue integrable functions are used but not when one considers only Riemann integrable functions. We shall investigate this situation more thoroughly later, especially in Chapter 6. [We have already proved the completeness of $L^1(\mu)$, disguised as Theorem (2.25). To remove the disguise, see Theorem 5.1 below.]

EXERCISES

19. If $\mu(X) < \infty$, $\{f_n\} \subset L^1(\mu)$, and $f_n \to f$ uniformly, then $f \in L^1(\mu)$ and $\int f_n \to \int f$.

20. If $f_n, g_n, f, g \in L^1$, $f_n \to f$ a.e., $|f_n| \leq g_n$, and $\int g_n \to \int g$, then $\int f_n \to \int f$. (Rework the proof of the dominated convergence theorem.)

21. Suppose $f_n, f \in L^1$ and $f_n \to f$ a.e. Then $\int |f_n - f| \to 0$ iff $\int |f_n| \to \int |f|$. (Use Exercise 20.)

22. Let μ be counting measure on $\mathbf{N}$. Interpret Fatou's lemma and the monotone and dominated convergence theorems as statements about infinite series.

23. Let $f\colon [a, b] \to \mathbf{R}$ be bounded, and let $H(x) = \limsup_{y\to x} f(y)$, $h(x) = \liminf_{y\to x} f(y)$.

(a) $H(x) = h(x)$ iff f is continuous at x.

(b) In the notation of the proof of Theorem (2.28a), $H = G$ a.e. and $h = g$ a.e. Hence H and h are Lebesgue measurable and $\int_{[a,b]} H\,dm = \overline{\int}_a^b f(x)\,dx$, $\int_{[a,b]} h\,dm = \underline{\int}_a^b f(x)\,dx$.

(c) Deduce Theorem (2.28b).

24. Let $(X, \mathcal{M}, \mu)$ be a measure space with $\mu(X) < \infty$, and let $(X, \overline{\mathcal{M}}, \bar{\mu})$ be its completion. If $f\colon X \to \mathbf{R}$ is bounded, the following are equivalent:

(a) f is $\overline{\mathcal{M}}$-measurable [and hence $f \in L^1(\bar{\mu})$].

(b) For each $n \in \mathbf{N}$ there exist $\mathcal{M}$-measurable simple functions φ_n and ψ_n with $\varphi_n \leq f \leq \psi_n$ and $\int(\psi_n - \varphi_n)\,d\mu < 1/n$.

If these conditions hold, $\lim \int \varphi_n\,d\mu = \lim \int \psi_n\,d\mu = \int f\,d\bar{\mu}$.

25. Let $f(x) = x^{-1/2}$ if $0 < x < 1$, $f(x) = 0$ otherwise. Let $\{r_n\}_1^\infty$ be an enumeration of the rationals, and set $g(x) = \sum_1^\infty 2^{-n} f(x - r_n)$.

(a) $g \in L^1(m)$, and in particular $g < \infty$ a.e.

(b) g is discontinuous at every point and unbounded on every interval, and it remains so after any modification on a Lebesgue null set.

(c) $g^2 < \infty$ a.e., but g^2 is not integrable on any interval.

26. If $f \in L^1(m)$ and $F(x) = \int_{-\infty}^x f(t)\,dt$, then F is continuous on $\mathbf{R}$.

27. If $f_n(x) = ae^{-nax} - be^{-nbx}$ where $0 < a < b$, then $\sum_1^\infty \int_0^\infty |f_n(x)|\,dx = \infty$, and $\sum_1^\infty \int_0^\infty f_n(x)\,dx \neq \int_0^\infty \sum_1^\infty f_n(x)\,dx$.

28. Compute the following limits and justify the calculations:

(a) $\lim_{n\to\infty} \int_0^\infty (1 + (x/n))^{-n} \sin(x/n)\,dx$

(b) $\lim_{n\to\infty} \int_0^1 (1 + nx^2)(1 + x^2)^{-n}\,dx$

(c) $\lim_{n\to\infty} \int_0^\infty n \sin(x/n)[x(1 + x^2)]^{-1}\,dx$

(d) $\lim_{n\to\infty} \int_a^\infty n(1 + n^2x^2)^{-1}\,dx$ (The answer depends on whether $a > 0$, $a = 0$, or $a < 0$. How does this accord with the various convergence theorems?)

29. (a) Show that $\int_0^\infty x^n e^{-x}\,dx = n!$ by differentiating the equation

$$\int_0^\infty e^{-tx}\,dt = 1/x.$$

(b) Similarly, show that $\int_{-\infty}^{\infty} x^{2n}e^{-x^2}\,dx = (2n)!/4^n n!\sqrt{\pi}$, assuming the fact that $\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$ [cf. Proposition (2.53) below].

30. Show that $\lim_{k\to\infty}\int_0^k x^n(1 - x/k)^k\,dx = n!$. (Use Exercise 29a.)

31. Derive the following formulas by expanding part of the integrand into an infinite series and justifying the term-by-term integration. Exercise 29 may be useful.

(a) $\int_{-\infty}^{\infty} e^{-x^2}\cos ax\,dx = \sqrt{\pi}\,e^{-a^2/4}$ $(a > 0)$.

(b) $\int_0^1 x^p(x - 1)^{-1}\,dx = \sum_1^\infty 1/(p + k)^2$ $(p > -1)$.

(c) $\int_0^\infty x^{a-1}(e^x - 1)^{-1}\,dx = \Gamma(a)\zeta(a)$ where $\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}\,dx$, $\zeta(a) = \sum_1^\infty n^{-a}$ $(a > 1)$.

(d) $\int_0^\infty e^{-sx}x^{-1}\sin x\,dx = \arctan(1/s)$ $(s > 1)$.

(e) $\int_0^\infty e^{-sx}J_0(x)\,dx = (s^2 + 1)^{-1/2}$, where $J_0(x) = \sum_0^\infty (-1)^n x^{2n}/4^n(n!)^2$ is the Bessel function of order zero $(s > 1)$.

[In (d) and (e), term-by-term integration works, and the resulting series converges, only for $s > 1$, but the formulas as stated are actually valid for all $s > 0$.]

4. MODES OF CONVERGENCE

If $\{f_n\}$ is a sequence of complex-valued functions on a set X, the statement "$f_n \to f$ as $n \to \infty$" can be taken in many different senses, for example, uniform or pointwise convergence. If X is a measure space, one can also speak of a.e. convergence or convergence in L^1 (i.e., $\int |f_n - f| \to 0$). Of course, uniform convergence implies pointwise convergence, which in turn implies a.e. convergence (and not conversely, in general), but these modes of convergence do not imply L^1 convergence or vice versa. It will be useful to keep in mind the following examples on $\mathbf{R}$ (with Lebesgue measure):

(1) $f_n = n^{-1}\chi_{(0,n)}$.
(2) $f_n = \chi_{(n,n+1)}$.
(3) $f_n = n\chi_{[0,1/n]}$.
(4) $f_n = \chi_{[j2^{-k},(j+1)2^{-k}]}$ where $0 \le j < 2^k$ and $n = j + 2^k$.

In (1), (2), and (3), $f_n \to 0$ uniformly, pointwise, and a.e., respectively, but $f_n \not\to 0$ in L^1 (in fact $\int |f_n| = 1$ for all n). In (4), $f_n \to 0$ in L^1 since $\int |f_n| = 2^{-k}$ for $2^k \le n < 2^{k+1}$; but $f_n(x)$ does not converge for any $x \in [0, 1]$ since there are infinitely many n for which $f_n(x) = 0$ and infinitely many for which $f_n(x) = 1$. (The reader should draw the graphs of the first few f_n's to see what is going on here.)

On the other hand, if $f_n \to f$ a.e. and $|f_n| \le g \in L^1$ for all n, then $f_n \to f$ in L^1. (This is clear from the dominated convergence theorem, since $|f_n - f| \le 2g$.) Also, we shall see below that if $f_n \to f$ in L^1 then some subsequence converges to f a.e.

Another mode of convergence that is frequently useful is convergence in measure. We say that a sequence $\{f_n\}$ of measurable complex-valued functions on $(X, \mathcal{M}, \mu)$ is **Cauchy in measure** if for every $\varepsilon > 0$,

$$\mu(\{x: |f_n(x) - f_m(x)| \ge \varepsilon\}) \to 0 \text{ as } m, n \to \infty,$$

and that $\{f_n\}$ **converges in measure** to f if for every $\varepsilon > 0$,

$$\mu(\{x: |f_n(x) - f(x)| \ge \varepsilon\}) \to 0 \text{ as } n \to \infty.$$

For example, the sequences (1), (3), and (4) above converge to zero in measure, but (2) is not Cauchy in measure.

(2.29) Proposition. If $f_n \to f$ in L^1, then $f_n \to f$ in measure.

Proof. Let $E_{n,\varepsilon} = \{x: |f_n(x) - f(x)| \ge \varepsilon\}$. Then

$$\int |f_n - f| \ge \int_{E_{n,\varepsilon}} |f_n - f| \ge \varepsilon\mu(E_{n,\varepsilon}),$$

so $\mu(E_{n,\varepsilon}) \le \varepsilon^{-1}\int |f_n - f| \to 0$. □

The converse of Proposition (2.29) is false, as examples (1) and (3) show.

(2.30) Theorem. Suppose $\{f_n\}$ is Cauchy in measure. Then there exists a measurable f such that $f_n \to f$ in measure, and there is a subsequence $\{f_{n_j}\}$ of $\{f_n\}$ which converges to f a.e. Moreover, if also $f_n \to g$ in measure, then $g = f$ a.e.

Proof. We can choose a subsequence $\{g_j\} = \{f_{n_j}\}$ of $\{f_n\}$ such that if $E_j = \{x: |g_j(x) - g_{j+1}(x)| \ge 2^{-j}\}$ then $\mu(E_j) \le 2^{-j}$. If $F_k = \bigcup_{j=k}^\infty E_j$, then $\mu(F_k) \le \sum_k^\infty 2^{-j} = 2^{1-k}$, and if $x \notin F_k$, for $i \ge j \ge k$ we have

$$|g_j(x) - g_i(x)| \le \sum_{l=j}^{i-1} |g_{l+1}(x) - g_l(x)| \le \sum_{l=j}^{i-1} 2^{-l} \le 2^{1-j}. \tag{2.31}$$

Thus $\{g_j\}$ is pointwise Cauchy on F_k^c. Let $F = \bigcap_1^\infty F_k = \limsup E_j$: then $\mu(F) = 0$, and if we set $f(x) = \lim g_j(x)$ for $x \notin F$ and $f(x) = 0$ for $x \in F$, then f is measurable (cf. Exercises 3 and 5) and $g_j \to f$ a.e. Also, (2.31) shows that $|g_j(x) - f(x)| \le 2^{2-j}$ for $j \ge k$, $x \notin F_k$. Since $\mu(F_k) \to 0$ as $k \to \infty$, it follows that $g_j \to f$ in measure. But then $f_n \to f$ in measure, because

$$\{x: |f_n(x) - f(x)| \ge \varepsilon\} \subset \{x: |f_n(x) - g_j(x)| \ge \varepsilon/2\} \cup \{x: |g_j(x) - f(x)| \ge \varepsilon/2\},$$

and the sets on the right both have small measure when n and j are large. Likewise, if $f_n \to g$ in measure.

$$\{x: |f(x) - g(x)| \geq \varepsilon\} \subset \{x: |f(x) - f_n(x)| \geq \varepsilon/2\} \cup \{x: |f_n(x) - g(x)| \geq \varepsilon/2\}$$

for all n, hence $\mu(\{x: |f(x) - g(x)| \geq \varepsilon\}) = 0$ for all ε. Letting ε tend to zero through some sequence of values, we conclude that $f = g$ a.e. □

(2.32) Corollary. If $f_n \to f$ in L^1, there is a subsequence $\{f_{n_j}\}$ such that $f_{n_j} \to f$ a.e.

Proof. Combine Proposition (2.29) and Theorem (2.30). □

If $f_n \to f$ a.e., it does not follow that $f_n \to f$ in measure, as example (2) shows. However, this conclusion does hold on a finite measure space, where in fact something considerably stronger is true.

(2.33) Egoroff's Theorem. Suppose $\mu(X) < \infty$, and f_n and f are measurable complex-valued functions on X such that $f_n \to f$ a.e. Then for every $\varepsilon > 0$ there exists $E \subset X$ such that $\mu(E) < \varepsilon$ and $f_n \to f$ uniformly on E^c.

Proof. Without loss of generality we may assume that $f_n \to f$ everywhere on X. For $k, n \in \mathbf{N}$ let

$$E_n(k) = \bigcup_{m=n}^{\infty} \{x: |f_m(x) - f(x)| \geq k^{-1}\}.$$

Then, for fixed k, $E_n(k)$ decreases as n increases, and $\bigcap_{n=1}^{\infty} E_n(k) = \varnothing$, so since $\mu(X) < \infty$ we conclude that $\mu(E_n(k)) \to 0$ as $n \to \infty$. Given $\varepsilon > 0$ and $k \in \mathbf{N}$, choose n_k so large that $\mu(E_{n_k}(k)) < \varepsilon 2^{-k}$ and let $E = \bigcup_{k=1}^{\infty} E_{n_k}(k)$. Then $\mu(E) < \varepsilon$, and we have $|f_n(x) - f(x)| < k^{-1}$ for $n > n_k$ and $x \notin E$. Thus $f_n \to f$ uniformly on E^c. □

The type of convergence involved in the conclusion of Egoroff's theorem is sometimes called **almost uniform convergence**. It is not hard to see that almost uniform convergence implies a.e. convergence and convergence in measure (Exercise 39).

EXERCISES

32. Suppose $\mu(X) < \infty$, and define

$$\rho(f) = \int \frac{|f|}{1 + |f|}\, d\mu.$$

Then the map $(f, g) \to \rho(f - g)$ is a metric on the space of complex-valued measurable functions on X (if we identify two functions which are equal a.e.), and $\rho(f_n - f) \to 0$ iff $f_n \to f$ in measure.

33. If $f_n \geq 0$ and $f_n \to f$ in measure, then $\int f \leq \liminf \int f_n$.

34. Suppose $|f_n| \leq g \in L^1$ and $f_n \to f$ in measure.

(a) $\int f = \lim \int f_n$.

(b) $f_n \to f$ in L^1.

35. $f_n \to f$ in measure iff for every $\varepsilon > 0$ there exists $N \in \mathbf{N}$ such that $\mu(\{x: |f_n(x) - f(x)| \geq \varepsilon\}) < \varepsilon$ for $n \geq N$.

36. If $\mu(E_n) < \infty$ for $n \in \mathbf{N}$ and $\chi_{E_n} \to f$ in L^1, then f is (a.e. equal to) the characteristic function of a measurable set.

37. Suppose f_n and f are measurable complex-valued functions and $\varphi: \mathbf{C} \to \mathbf{C}$.

(a) If φ is continuous and $f_n \to f$ a.e., then $\varphi \circ f_n \to \varphi \circ f$ a.e.

(b) If φ is uniformly continuous and $f_n \to f$ uniformly, almost uniformly, or in measure, then $\varphi \circ f_n \to \varphi \circ f$ uniformly, almost uniformly, or in measure, respectively.

(c) There are counterexamples when the continuity assumptions on φ are not satisfied.

38. Suppose $f_n \to f$ in measure and $g_n \to g$ in measure.

(a) $f_n + g_n \to f + g$ in measure.

(b) $f_n g_n \to fg$ in measure if $\mu(X) < \infty$ but not necessarily if $\mu(X) = \infty$.

39. If $f_n \to f$ almost uniformly, then $f_n \to f$ a.e. and in measure.

40. In Egoroff's theorem the hypothesis "$\mu(X) < \infty$" can be replaced by "$|f_n| \leq g$ for all n, where $g \in L^1(\mu)$."

41. If μ is σ-finite and $f_n \to f$ a.e., there exist measurable $E_1, E_2, \ldots \subset X$ such that $\mu((\bigcup_1^{\infty} E_j)^c) = 0$ and $f_n \to f$ uniformly on each E_j.

42. Let μ be counting measure on $\mathbf{N}$. Then $f_n \to f$ in measure iff $f_n \to f$ uniformly.

43. Suppose $\mu(X) < \infty$ and $f: X \times [0, 1] \to \mathbf{C}$ is a function such that $f(\cdot, y)$ is measurable for each $y \in [0, 1]$ and $f(x, \cdot)$ is continuous for each $x \in X$.

(a) If $0 < \varepsilon, \delta < 1$ then $E_{\varepsilon,\delta} = \{x: |f(x, y) - f(x, 0)| \geq \varepsilon \text{ for all } y < \delta\}$ is measurable.

(b) For any $\varepsilon > 0$ there is a set $E \subset X$ such that $\mu(E) < \varepsilon$ and $f(\cdot, y) \to f(\cdot, 0)$ uniformly on E^c as $y \to 0$.

44. (Lusin's Theorem) If $f: [a, b] \to \mathbf{C}$ is Lebesgue measurable and $\varepsilon > 0$, there is a set $E \subset [a, b]$ such that $m(E^c) < \varepsilon$ and $f|E$ is continuous. Moreover, E may be taken to be compact. [Use Egoroff's theorem and Theorem (2.26).]

5. PRODUCT MEASURES

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be measure spaces. We have already discussed the product σ-algebra $\mathcal{M} \otimes \mathcal{N}$ on $X \times Y$; we now construct a measure on $\mathcal{M} \otimes \mathcal{N}$ which is, in an obvious sense, the product of μ and ν.

To begin with, we define a **rectangle** to be a set of the form $A \times E$ where $A \in \mathcal{M}$ and $E \in \mathcal{N}$. Clearly $(A \times E) \cap (B \times F) = (A \cap B) \times (E \cap F)$, and $(A \times E)^c = (X \times E^c) \cup (A^c \times E)$. Therefore, by Proposition (1.7), the collection $\mathcal{A}$ of finite disjoint unions of rectangles is an algebra, and of course $\mathcal{M} \otimes \mathcal{N}$ is the σ-algebra generated by $\mathcal{A}$.

Suppose $A \times E$ is a rectangle which is a (finite or countable) disjoint union of rectangles $A_j \times E_j$. Then for $x \in X$ and $y \in Y$,

$$\chi_A(x)\chi_E(y) = \chi_{A\times E}(x, y) = \sum \chi_{A_j \times E_j}(x, y) = \sum \chi_{A_j}(x)\chi_{E_j}(y).$$

If we integrate with respect to x and use Theorem (2.15), we obtain

$$\mu(A)\chi_E(y) = \chi_E(y) \int \chi_A(x)\, d\mu(x)$$

$$= \sum \chi_{E_j}(y) \int \chi_{A_j}(x)\, d\mu(x) = \sum \mu(A_j)\chi_{E_j}(y).$$

In the same way, integration in y then yields

$$\mu(A)\nu(E) = \sum \mu(A_j)\nu(E_j).$$

It follows that if $B \in \mathcal{A}$ is the disjoint union of rectangles $A_1 \times E_1, \ldots, A_n \times E_n$ and we set

$$\pi(B) = \sum_1^n \mu(A_j)\nu(E_j)$$

(with the usual convention that $0 \cdot \infty = 0$), then π is well defined on $\mathcal{A}$ (since any two representations of B as a finite disjoint union of rectangles have a common refinement), and π is a premeasure on $\mathcal{A}$. According to Theorem (1.14), therefore, π generates an outer measure on $X \times Y$ whose restriction to $\mathcal{M} \otimes \mathcal{N}$ is a measure which extends π. We call this measure the **product** of μ and ν and denote it by $\mu \times \nu$. Moreover, if μ and ν are σ-finite—say, $X = \bigcup_1^\infty A_j$ and $Y = \bigcup_1^\infty E_k$ with $\mu(A_j) < \infty$ and $\nu(E_k) < \infty$—then $X \times Y = \bigcup_{j,k} A_j \times E_k$, and $\mu \times \nu(A_j \times E_k) < \infty$, so $\mu \times \nu$ is σ-finite. $\mu \times \nu$ is then the unique measure on $\mathcal{M} \otimes \mathcal{N}$ such that $\mu \times \nu(A \times E) = \mu(A)\nu(E)$ for all rectangles $A \times E$.

The same construction works for any finite number of factors. That is, suppose $(X_j, \mathcal{M}_j, \mu_j)$ $(j = 1, \ldots, n)$ are measure spaces. If we define a rectangle to be a set of the form $A_1 \times \cdots \times A_n$ with $A_j \in \mathcal{M}_j$, then the collection $\mathcal{A}$ of finite disjoint unions of rectangles is an algebra, and the same procedure as above produces a measure $\mu_1 \times \cdots \times \mu_n$ on $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ such that

$$\mu_1 \times \cdots \times \mu_n(A_1 \times \cdots \times A_n) = \prod_1^n \mu_j(A_j).$$

Moreover, if the μ_j's are σ-finite so that the product is uniquely determined, the obvious associativity properties hold. For example, if we identify $X_1 \times X_2 \times X_3$ with $(X_1 \times X_2) \times X_3$, we have $\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3 = (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3$ (the former being generated by sets of the form $A_1 \times A_2 \times A_3$ with $A_j \in \mathcal{M}_j$, and the latter by sets of the form $B \times A_3$ with $B \in \mathcal{M}_1 \otimes \mathcal{M}_2$, $A_3 \in \mathcal{M}_3$), and $\mu_1 \times \mu_2 \times \mu_3 = (\mu_1 \times \mu_2) \times \mu_3$ (since they agree on sets of the form $A_1 \times A_2 \times A_3$, hence in general by uniqueness). Details are left to the reader (Exercise 45). All of our results below have obvious extensions to products with n factors, but we shall stick to the case $n = 2$ for simplicity.

We return to the case of two measure spaces $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$. If $E \subset X \times Y$, for $x \in X$ and $y \in Y$ we define the ***x*-section** E_x and the ***y*-section** E^y of E by

$$E_x = \{y \in Y: (x, y) \in E\}, \qquad E^y = \{x \in X: (x, y) \in E\}.$$

Also, if f is a function on $X \times Y$ we define the ***x*-section** f_x and the ***y*-section** f^y of f by

$$f_x(y) = f^y(x) = f(x, y).$$

(2.34) Proposition.

(a) If $E \in \mathcal{M} \otimes \mathcal{N}$, then $E_x \in \mathcal{N}$ for all $x \in X$ and $E^y \in \mathcal{M}$ for all $y \in Y$.

(b) If f is $\mathcal{M} \otimes \mathcal{N}$-measurable, then f_x is $\mathcal{N}$-measurable for all $x \in X$ and f^y is $\mathcal{M}$-measurable for all $y \in Y$.

Proof. (a) Let $\mathcal{R} = \{E \subset X \times Y: E_x \in \mathcal{N}$ for all x and $E^y \in \mathcal{M}$ for all $y\}$. Then $\mathcal{R}$ obviously contains all rectangles [e.g., $(A \times B)_x = B$ if $x \in A$, $= \varnothing$ otherwise]. Since $(\bigcup_1^\infty E_j)_x = \bigcup_1^\infty (E_j)_x$ and $(E_x)^c = (E^c)_x$, $\mathcal{R}$ is a σ-algebra. Therefore, $\mathcal{R} \supset \mathcal{M} \otimes \mathcal{N}$.

(b) follows from (a) because $(f_x)^{-1}(B) = (f^{-1}(B))_x$ and $(f^y)^{-1}(B) = (f^{-1}(B))^y$. □

Remark. Even if μ and ν are complete, $\mu \times \nu$ is almost never complete. Indeed, suppose there is a nonempty $A \in \mathcal{M}$ with $\mu(A) = 0$ and that $\mathcal{N} \neq \mathcal{P}(Y)$. (This is the case with $\mu = \nu =$ Lebesgue measure on $\mathbf{R}$, for instance.) If

$E \in \mathscr{P}(Y) \setminus \mathscr{N}$, then $A \times E \notin \mathscr{M} \otimes \mathscr{N}$ by Proposition (2.34), but $A \times E \subset A \times Y$ and $\mu \times \nu(A \times Y) = 0$.

Before proceeding further we need a technical lemma. We define a **monotone class** on a space X to be a subset $\mathscr{C}$ of $\mathscr{P}(X)$ which is closed under countable increasing unions and countable decreasing intersections (i.e., if $E_j \in \mathscr{C}$ and $E_1 \subset E_2 \subset \cdots$, then $\bigcup E_j \in \mathscr{C}$, and likewise for intersections). Clearly every σ-algebra is a monotone class. Also, the intersection of any family of monotone classes is a monotone class, so for any $\mathscr{E} \subset \mathscr{P}(X)$ there is a unique smallest monotone class containing $\mathscr{E}$, called the monotone class **generated by** $\mathscr{E}$.

(2.35) Lemma. If $\mathscr{A}$ is an algebra of sets, then the monotone class $\mathscr{C}$ generated by $\mathscr{A}$ coincides with the σ-algebra $\mathscr{M}$ generated by $\mathscr{A}$.

Proof. Since $\mathscr{M}$ is a monotone class we have $\mathscr{C} \subset \mathscr{M}$. If $E \in \mathscr{C}$, let

$$\mathscr{C}(E) = \{F \in \mathscr{C}\colon E \setminus F,\ F \setminus E, \text{ and } E \cap F \text{ are in } \mathscr{C}\}.$$

Clearly $\varnothing, E \in \mathscr{C}(E)$, and $E \in \mathscr{C}(F)$ iff $F \in \mathscr{C}(E)$. Also, $\mathscr{C}(E)$ is easily seen to be a monotone class. If $E \in \mathscr{A}$ then $F \in \mathscr{C}(E)$ for all $F \in \mathscr{A}$ because $\mathscr{A}$ is an algebra; that is, $\mathscr{A} \subset \mathscr{C}(E)$, and hence $\mathscr{C} \subset \mathscr{C}(E)$. Therefore, if $F \in \mathscr{C}$ then $F \in \mathscr{C}(E)$ for all $E \in \mathscr{A}$. But this means that $E \in \mathscr{C}(F)$ for all $E \in \mathscr{A}$, so that $\mathscr{A} \subset \mathscr{C}(F)$ and hence $\mathscr{C} \subset \mathscr{C}(F)$. Conclusion: if $E, F \in \mathscr{C}$, then $E \setminus F$ and $E \cap F$ are in $\mathscr{C}$. Since $X \in \mathscr{A} \subset \mathscr{C}$, $\mathscr{C}$ is therefore an algebra. If $\{E_j\}_1^\infty \subset \mathscr{C}$, then, $F_n = \bigcup_1^n E_j \in \mathscr{C}$ for all n, and the closure of $\mathscr{C}$ under increasing unions then implies that $\bigcup_1^\infty E_j = \bigcup_1^\infty F_n \in \mathscr{C}$. In short, $\mathscr{C}$ is a σ-algebra, so $\mathscr{C} \supset \mathscr{M}$. □

We now come to the main results of this section, which relate integrals on $X \times Y$ to integrals on X and Y.

(2.36) Theorem. Suppose $(X, \mathscr{M}, \mu)$ and $(Y, \mathscr{N}, \nu)$ are σ-finite measure spaces. If $E \in \mathscr{M} \otimes \mathscr{N}$, then the functions $x \to \nu(E_x)$ and $y \to \mu(E^y)$ are measurable on X and Y, respectively, and

$$\mu \times \nu(E) = \int \nu(E_x)\, d\mu\,(x) = \int \mu(E^y)\, d\nu\,(y).$$

Proof. First suppose that μ and ν are finite, and let $\mathscr{C}$ be the set of all $E \in \mathscr{M} \otimes \mathscr{N}$ for which the conclusions of the theorem are true. If $E = A \times B$, then $\nu(E_x) = \chi_A(x)\nu(B)$ and $\mu(E^y) = \mu(A)\chi_B(y)$, so clearly $E \in \mathscr{C}$. By additivity it follows that finite disjoint unions of rectangles are in $\mathscr{C}$, so by Lemma (2.35) it will suffice to show that $\mathscr{C}$ is a monotone class. If $\{E_n\}$ is an increasing sequence in $\mathscr{C}$ and $E = \bigcup_1^\infty E_n$, then the functions $f_n(y) = \mu((E_n)^y)$ are measurable and increase pointwise to $f(y) = \mu(E^y)$. Hence f is measurable, and by the monotone convergence theorem,

$$\int \mu(E^y)\, d\nu\,(y) = \lim \int \mu\big((E_n)^y\big)\, d\nu\,(y) = \lim \mu \times \nu(E_n) = \mu \times \nu(E).$$

Likewise $\mu \times \nu(E) = \int \nu(E_x)\, d\mu\,(x)$, so $E \in \mathscr{C}$. Similarly, if $\{E_n\}$ is a decreasing sequence and $E = \bigcap_1^\infty E_n$, the function $y \to \mu((E_1)^y)$ is in $L^1(\nu)$ because $\mu((E_1)^y) \le \mu(X) < \infty$ and $\nu(Y) < \infty$, so the dominated convergence theorem can be applied to show that $E \in \mathscr{C}$. This completes the proof for the case of finite measure spaces.

Finally, if μ and ν are σ-finite we can write $X \times Y$ as the union of an increasing sequence $\{X_j \times Y_j\}$ of rectangles of finite measure. If $E \in \mathscr{M} \otimes \mathscr{N}$, the preceding argument applies to $E \cap (X_j \times Y_j)$ for each j to give

$$\mu \times \nu\big(E \cap (X_j \times Y_j)\big) = \int \chi_{X_j}(x)\nu\big(E_x \cap Y_j\big)\, d\mu\,(x)$$

$$= \int \chi_{Y_j}(y)\mu\big(E^y \cap X_j\big)\, d\nu\,(y),$$

and a final application of the monotone convergence theorem then yields the desired result. □

(2.37) The Fubini–Tonelli Theorem. Suppose $(X, \mathscr{M}, \mu)$ and $(Y, \mathscr{N}, \nu)$ are σ-finite measure spaces.

(a) (Tonelli) If $f \in L^+(X \times Y)$, then the functions $g(x) = \int f_x\, d\nu$ and $h(y) = \int f^y\, d\mu$ are in $L^+(X)$ and $L^+(Y)$, respectively, and

$$\int f d(\mu \times \nu) = \int \left[\int f(x, y)\, d\nu\,(y)\right] d\mu\,(x) \tag{2.38}$$

$$= \int \left[\int f(x, y)\, d\mu\,(x)\right] d\nu\,(y).$$

(b) (Fubini) If $f \in L^1(\mu \times \nu)$, then $f_x \in L^1(\nu)$ for a.e. $x \in X$, $f^y \in L^1(\mu)$ for a.e. $y \in Y$, the a.e.-defined functions $g(x) = \int f_x\, d\nu$ and $h(y) = \int f^y\, d\mu$ are in $L^1(\mu)$ and $L^1(\nu)$, respectively, and (2.38) holds.

Proof. Tonelli's theorem reduces to Theorem (2.36) in case f is a characteristic function, and it therefore holds for nonnegative simple functions by linearity. If $f \in L^+(X \times Y)$, let $\{f_n\}$ be a sequence of simple functions which increase pointwise to f as in Theorem (2.10). The monotone convergence theorem implies, first, that the corresponding g_n and h_n increase to g and h,

respectively (so that g and h are measurable), and, second, that

$$\int g\,d\mu = \lim \int g_n\,d\mu = \lim \int f_n d(\mu \times \nu) = \int f d(\mu \times \nu),$$

$$\int h\,d\nu = \lim \int h_n\,d\nu = \lim \int f_n d(\mu \times \nu) = \int f d(\mu \times \nu),$$

which is (2.38). This establishes Tonelli's theorem and also shows that if $f \in L^+(X \times Y)$ and $\int f d(\mu \times \nu) < \infty$, then $g < \infty$ a.e. and $h < \infty$ a.e., that is, $f_x \in L^1(\nu)$ for a.e. x and $f^y \in L^1(\mu)$ for a.e. y. If $f \in L^1(\mu \times \nu)$, then, the conclusion of Fubini's theorem follows by applying these results to the positive and negative parts of the real and imaginary parts of f. □

Remarks. (i) We shall usually omit writing the brackets in the iterated integrals in (2.38), thus:

$$\int \left[\int f(x, y)\,d\mu\,(x) \right] d\nu\,(y) = \iint f(x, y)\,d\mu\,(x)\,d\nu\,(y) = \iint f\,d\mu\,d\nu.$$

(ii) The hypothesis of σ-finiteness is necessary: see Exercise 46.

(iii) The hypothesis $f \in L^+(X \times Y)$ or $f \in L^1(\mu \times \nu)$ is necessary, in two respects. First, it is possible for f_x and f^y to be measurable for all x, y and for the iterated integrals $\int\int f\,d\mu\,d\nu$ and $\int\int f\,d\nu\,d\mu$ to exist even if f is not $\mathcal{M} \otimes \mathcal{N}$-measurable. However, the iterated integrals need not then be equal: see Exercise 47. Second, if f is not nonnegative, it is possible for f_x and f^y to be integrable for all x, y and for the iterated integrals $\int\int f\,d\mu\,d\nu$ and $\int\int f\,d\nu\,d\mu$ to exist even if $\int |f|\,d(\mu \times \nu) = \infty$. But, again, the iterated integrals need not be equal: see Exercise 48.

(iv) The Fubini and Tonelli theorems are frequently used in tandem. Typically one wishes to invert the order of integration in a double integral $\int\int f\,d\mu\,d\nu$. *First* one verifies that $\int |f|\,d(\mu \times \nu) < \infty$ by using Tonelli's theorem to evaluate this integral as an iterated integral; *then* one applies Fubini's theorem to conclude that $\int\int f\,d\mu\,d\nu = \int\int f\,d\nu\,d\mu$. (For examples, see the exercises in Section 6.)

We conclude with another version of the Fubini–Tonelli theorem which deals with complete measures; this is relevant to Lebesgue integration on $\mathbf{R}^n$.

(2.39) Theorem. Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are complete, σ-finite measure spaces, and let $(X \times Y, \mathcal{L}, \lambda)$ be the completion of $(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$. If f is $\mathcal{L}$-measurable and either (a) $f \geq 0$ or (b) $f \in L^1(\lambda)$, then f_x is $\mathcal{N}$-measurable for a.e. x and f^y is $\mathcal{M}$-measurable for a.e. y, and in case (b) f_x and f^y are also integrable for a.e. x and y. Moreover, $x \to \int f_x\,d\nu$ and $y \to \int f^y\,d\mu$ are measurable, and in case (b) also integrable, and

$$\int f\,d\lambda = \iint f(x, y)\,d\mu\,(x)\,d\nu\,(y) = \iint f(x, y)\,d\nu\,(y)\,d\mu\,(x).$$

The proof is outlined in Exercise 49.

EXERCISES

45. If $(X_j, \mathcal{M}_j)$ is a measurable space for $j = 1, 2, 3$, then $\otimes_1^3 \mathcal{M}_j = (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3$. Moreover, if μ_j is a σ-finite measure on $(X_j, \mathcal{M}_j)$, then $\mu_1 \times \mu_2 \times \mu_3 = (\mu_1 \times \mu_2) \times \mu_3$.

46. Let $X = Y = [0, 1]$, $\mathcal{M} = \mathcal{N} = \mathcal{B}_{[0,1]}$, μ = Lebesgue measure, and ν = counting measure. If $D = \{(x, y) \in X \times Y\colon x = y\}$, then $\int\int \chi_D\,d\mu\,d\nu$, $\int\int \chi_D\,d\nu\,d\mu$, and $\int \chi_D d(\mu \times \nu)$ are all unequal. [To compute $\int \chi_D d(\mu \times \nu) = \mu \times \nu(D)$, go back to the definition of $\mu \times \nu$.]

47. Let $X = Y$ be an uncountable linearly ordered set such that for each $x \in X$, $\{y\colon y < x\}$ is countable. (Example: the set of countable ordinals.) Let $\mathcal{M} = \mathcal{N}$ be the σ-algebra of countable or co-countable sets, and let $\mu = \nu$ be defined on $\mathcal{M}$ by $\mu(A) = 0$ if A is countable, $\mu(A) = 1$ otherwise. Let $E = \{(x, y)\colon y < x\}$. Then E_x and E^y are measurable for all x, y, and $\int\int \chi_E\,d\mu\,d\nu$ and $\int\int \chi_E\,d\nu\,d\mu$ exist and are unequal. (Hence $E \notin \mathcal{M} \otimes \mathcal{N}$. If one believes in the continuum hypothesis, one can take $X = [0, 1]$ [with a nonstandard ordering], and thus obtain a set $E \subset [0, 1] \times [0, 1]$ such that E_x is countable and E^y is co-countable [in particular, Borel] for all x, y, but E is not Lebesgue measurable.)

48. Let $X = Y = \mathbf{N}$, $\mathcal{M} = \mathcal{N} = \mathcal{P}(\mathbf{N})$, $\mu = \nu$ = counting measure. Define $f(m, n) = 1$ if $m = n$, $f(m, n) = -1$ if $m = n + 1$, and $f(m, n) = 0$ otherwise. Then $\int |f|\,d(\mu \times \nu) = \infty$, and $\int\int f\,d\mu\,d\nu$ and $\int\int f\,d\nu\,d\mu$ exist and are unequal.

49. Prove Theorem (2.39) as follows:

(a) If $E \in \mathcal{M} \otimes \mathcal{N}$ and $\mu \times \nu(E) = 0$, then $\nu(E_x) = \mu(E^y) = 0$ for a.e. x and y.

(b) If f is $\mathcal{L}$-measurable and $f = 0$ λ-a.e., then f_x and f^y are integrable for a.e. x and y, and $\int f_x\,d\nu = \int f^y\,d\mu = 0$ for a.e. x and y. (Here the completeness of μ and ν is needed.)

(c) Now use Proposition (2.12).

50. Suppose $(X, \mathcal{M}, \mu)$ is a σ-finite measure space, and $f \in L^+(X)$. Let $G_f = \{(x, y) \in X \times [0, \infty]\colon 0 \leq y \leq f(x)\}$. Then G_f is $\mathcal{M} \otimes \mathcal{B}_{\mathbf{R}}$-measurable and $\mu \times m(G_f) = \int f\,d\mu$. [Hint: the map $(x, y) \to (f(x), y)$ is $(\mathcal{M} \otimes \mathcal{B}_{\mathbf{R}}, \mathcal{B}_{\mathbf{R}^2})$-measurable and $(z, y) \to z - y$ is continuous on $\mathbf{R}^2$.] This is the definitive statement of the familiar theorem from calculus, "the integral of a function is the area under its graph."

51. Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be arbitrary measure spaces (not necessarily σ-finite).

(a) If $f\colon X \to \mathbf{C}$ is $\mathcal{M}$-measurable, $g\colon Y \to \mathbf{C}$ is $\mathcal{N}$-measurable, and h is defined on $X \times Y$ by $h(x, y) = f(x)g(y)$, then h is $\mathcal{M} \otimes \mathcal{N}$-measurable.

(b) If $f \in L^1(\mu)$ and $g \in L^1(\nu)$, then $h \in L^1(\mu \times \nu)$ and $\int h d(\mu \times \nu) = \int f d\mu \int g d\nu$.

52. The Fubini–Tonelli theorem holds for an arbitrary measure space $(X, \mathcal{M}, \mu)$ when Y is a countable set, $\mathcal{N} = \mathcal{P}(Y)$, and ν is counting measure. [Cf. Theorems (2.15) and (2.25).]

6. THE LEBESGUE INTEGRAL ON $\mathbf{R}^n$

Lebesgue measure m^n on $\mathbf{R}^n$ is the completion of the n-fold product of Lebesgue measure on $\mathbf{R}$ with itself, that is, the completion of $m \times \cdots \times m$ on $\mathcal{B}_{\mathbf{R}} \otimes \cdots \otimes \mathcal{B}_{\mathbf{R}} = \mathcal{B}_{\mathbf{R}^n}$, or equivalently the completion of $m \times \cdots \times m$ on $\mathcal{L} \otimes \cdots \otimes \mathcal{L}$. The domain $\mathcal{L}^n$ of m^n is the class of **Lebesgue measurable sets** in $\mathbf{R}^n$; sometimes we shall also consider m^n as a measure on the smaller domain $\mathcal{B}_{\mathbf{R}^n}$. When there is no danger of confusion, we shall usually omit the superscript n and write m for m^n, and as in the case $n = 1$, we shall usually write $\int f(x)\,dx$ for $\int f dm^n$.

Most of the results of Section 1.5 have analogs for n-dimensional Lebesgue measure. In what follows, if $E = \prod_1^n E_j$ is a rectangle in $\mathbf{R}^n$, we shall refer to the sets $E_j \subset \mathbf{R}$ as the **sides** of E.

(2.40) Theorem. Suppose $E \in \mathcal{L}^n$.

(a) $m(E) = \inf\{m(U)\colon U \supset E,\ U \text{ open}\} = \sup\{m(K)\colon K \subset E,\ K \text{ compact}\}$.

(b) $E = A_1 \cup N_1 = A_2 \setminus N_2$ where A_1 is an F_σ set, A_2 is a G_δ set, and $m(N_1) = m(N_2) = 0$.

(c) If $m(E) < \infty$, for any $\varepsilon > 0$ there is a finite collection $\{R_j\}_1^N$ of disjoint rectangles whose sides are intervals such that $m(E \Delta \bigcup_1^N R_j) < \varepsilon$.

Proof. By the definition of product measures, if $E \in \mathcal{L}^n$ and $\varepsilon > 0$ there is a sequence $\{R_j\}$ of rectangles such that $E \subset \bigcup_1^\infty R_j$ and $\sum_1^\infty m(R_j) \le m(E) + \varepsilon$. For each j, by applying Theorem (1.18) to the sides of R_j we can find a rectangle $S_j \supset R_j$ whose sides are open sets such that $m(S_j) \le m(R_j) + \varepsilon 2^{-j}$. If $U = \bigcup_1^\infty S_j$, then U is open and $m(U) \le \sum_1^\infty m(S_j) \le m(E) + 2\varepsilon$. This proves the first equation in (a); the second one, and part (b), then follow as in the proofs of Theorems (1.18) and (1.19). Next, if $m(E) < \infty$ then $m(S_j) < \infty$ for all j; since the sides of S_j are countable unions of open intervals, by taking suitable finite subunions we obtain rectangles $T_j \subset S_j$ whose sides are finite unions of intervals such that $m(T_j) \ge m(S_j) - \varepsilon 2^{-j}$. If N is sufficiently large,

then, we have

$$m\left(E \Delta \bigcup_1^N T_j\right) = m\left(E \setminus \bigcup_1^N T_j\right) + m\left(\bigcup_1^N T_j \setminus E\right)$$

$$\le m\left(\bigcup_1^N (S_j \setminus T_j)\right) + m\left(\bigcup_{N+1}^\infty S_j\right) + m\left(\bigcup_1^\infty S_j \setminus E\right)$$

$$\le 3\varepsilon.$$

Since $\bigcup_1^N T_j$ can be expressed as a finite disjoint union of rectangles whose sides are intervals, we have proved (c). □

(2.41) Theorem. If $f \in L^1(m)$ and $\varepsilon > 0$, there exists $\varphi = \sum_1^N a_j \chi_{R_j}$ where each R_j is a product of intervals such that $\int |f - \varphi| < \varepsilon$, and there exists a continuous function g which vanishes outside a bounded set such that $\int |f - g| < \varepsilon$.

Proof. As in the proof of Theorem (2.26), approximate f by simple functions, then use Theorem (2.40c) to approximate the latter by functions φ of the desired form. Finally, approximate such φ's by continuous functions by an obvious generalization of the argument in Theorem (2.26). □

(2.42) Theorem. Lebesgue measure is translation invariant. That is, if $E \in \mathcal{L}^n$, $x \in \mathbf{R}^n$, and $E + x = \{y + x\colon y \in E\}$, then $E + x \in \mathcal{L}^n$ and $m(E + x) = m(E)$. Moreover, if f is Lebesgue measurable and either $f \ge 0$ or $f \in L^1(m)$, then $\int f(x + y)\,dy = \int f(y)\,dy$.

Proof. The first assertion follows easily from the one-dimensional result [Theorem (1.21)] if E is a rectangle, and by uniqueness of extensions [Theorem (1.14)] it then follows if E is a Borel set. In particular, the collection of Borel null sets is translation invariant, and this implies the result for arbitrary $E \in \mathcal{L}^n$. The second assertion reduces to the first when f is a characteristic function. By linearity it is then true for simple functions, so by the definition of the integral it is true for nonnegative measurable functions. Taking positive and negative parts of real and imaginary parts thus yields the result for $f \in L^1(m)$. □

We now investigate the behavior of the Lebesgue integral under linear transformations. We identify a linear map $T\colon \mathbf{R}^n \to \mathbf{R}^n$ with the matrix $(T_{ij}) = ((Te_j) \cdot e_i)$ where $\{e_j\}$ is the standard basis for $\mathbf{R}^n$, and we denote by $\det T$ the determinant of this matrix: thus $\det(T \circ S) = (\det T)(\det S)$. Furthermore, we employ the standard notation $GL(n, \mathbf{R})$ for the group of invertible linear transformations of $\mathbf{R}^n$. We shall need the fact from elementary linear algebra that every $T \in GL(n, \mathbf{R})$ can be written as the product of finitely many

transformations of the following three types:

$$T_1(x_1,\dots,x_j,\dots,x_n) = (x_1,\dots,cx_j,\dots,x_n) \quad (1 \le j \le n, c \ne 0),$$

$$T_2(x_1,\dots,x_j,\dots,x_n) = (x_1,\dots,x_j + cx_k,\dots,x_n)$$

$$(1 \le j \le n, k \ne j, c \ne 0),$$

$$T_3(x_1,\dots,x_j,\dots,x_k,\dots,x_n) = (x_1,\dots,x_k,\dots,x_j,\dots,x_n)$$

$$(1 \le j < k \le n).$$

(This is just a restatement of the fact that every nonsingular $n \times n$ matrix can be row-reduced to the identity matrix.)

(2.43) Theorem. Suppose $T \in GL(n, \mathbf{R})$.

(a) If f is a Lebesgue measurable function on $\mathbf{R}^n$, so is $f \circ T$. If $f \ge 0$ or $f \in L^1(m)$, then

$$\int f(x)\,dx = |\det T| \int f \circ T(x)\,dx. \tag{2.44}$$

(b) If $E \in \mathscr{L}^n$, then $T(E) \in \mathscr{L}^n$ and $m(T(E)) = |\det T| m(E)$.

Proof. If f is Borel measurable, so is $f \circ T$ since T is continuous. If (2.44) is true for $T, S \in GL(n, \mathbf{R})$ it is true for $T \circ S$, since

$$\int f(x)\,dx = |\det T| \int f \circ T(x)\,dx = |\det T|\,|\det S| \int f \circ T \circ S(x)\,dx$$

$$= |\det(T \circ S)| \int f \circ T \circ S(x)\,dx.$$

Hence it suffices to prove (2.44) when T is of the types T_1, T_2, T_3 described above. But this is a simple consequence of the Fubini–Tonelli theorem. For T_3 we merely interchange the order of integration in the variables x_j and x_k, and for T_1 and T_2 we integrate first with respect to x_j and use the one-dimensional formulas

$$\int f(t)\,dt = |c| \int f(ct)\,dt, \qquad \int f(t + a)\,dt = \int f(t)\,dt,$$

which follow from Theorem (1.21). Since it is easily verified that $\det T_1 = c$, $\det T_2 = 1$, and $\det T_3 = -1$, (2.44) is proved. Moreover, if E is a Borel set, so

is $T(E)$ (since T^{-1} is continuous), and by taking $f = \chi_{T(E)}$ we obtain $m(T(E)) = |\det T| m(E)$. In particular, the class of Borel null sets is invariant under T and T^{-1}, and hence so is $\mathscr{L}^n$. The result for Lebesgue measurable functions and sets now follows easily. □

(2.45) Corollary. Lebesgue measure is rotation-invariant.

Proof. Rotations are linear maps satisfying $TT^* = I$ where T^* is the transpose of T. Since $\det T = \det T^*$, this implies that $|\det T| = 1$. □

Next we shall generalize Theorem (2.43) from linear maps to differentiable maps. (This result will not be used elsewhere in this book and may be omitted on a first reading. We shall prove a generalization of it, by somewhat different methods, in Section 10.2.) To accomplish this, we need some terminology and a technical lemma.

Let $G = (g_1,\dots,g_n)$ be a map from an open set $\Omega \subset \mathbf{R}^n$ into $\mathbf{R}^n$ whose components g_j are of class C^1, that is, have continuous first-order partial derivatives. We denote by D_xG the linear map defined by the matrix $((\partial g_i/\partial x_j)(x))$ of partial derivatives at x. (Observe that if G is linear, then $D_xG = G$ for all x.) G is called a C^1 **diffeomorphism** if G is injective and D_xG is invertible for all $x \in \Omega$. In this case, the inverse function theorem guarantees that $G^{-1}\colon G(\Omega) \to \Omega$ is also a C^1 diffeomorphism and that $D_xG^{-1} = (D_{G^{-1}(x)}G)^{-1}$ for all $x \in G(\Omega)$.

By a **cube** we shall mean a product of compact intervals of equal lengths, that is, a set of the form $\{x\colon a_j \le x_j \le a_j + h\}$ for some $a \in \mathbf{R}^n$ and $h > 0$. Two cubes will be called **almost disjoint** if their interiors are disjoint. Since the boundary of a cube is contained in a finite union of hyperplanes, it has Lebesgue measure zero, so almost disjoint cubes may be considered as disjoint for the purposes of measure theory.

In what follows it will be convenient to use the norm

$$\|x\| = \max_{1 \le j \le n} |x_j|$$

on $\mathbf{R}^n$ rather than the Euclidean norm. Thus $\{x\colon \|x - a\| \le h\}$ is the cube of side length $2h$ centered at a. Moreover, if T is a linear transformation on $\mathbf{R}^n$ defined by the matrix (T_{ij}), we set

$$\|T\| = \max_{1 \le i \le n} \sum_{j=1}^{n} |T_{ij}|.$$

We then have $\|Tx\| \le \|T\|\,\|x\|$ for all $x \in \mathbf{R}^n$.

(2.46) Lemma. Every open set $U \subset \mathbf{R}^n$ is a countable almost disjoint union of cubes.

Proof. For $k \in \mathbf{N}$, let $\mathcal{Q}_k$ be the collection of cubes of side length 2^{-k} whose vertices have coordinates of the form $m2^{-k}$, $m \in \mathbf{Z}$. (Thus the cubes in $\mathcal{Q}_k$ are obtained from those in $\mathcal{Q}_{k-1}$ by bisecting the sides.) Two cubes in $\bigcup_1^\infty \mathcal{Q}_k$ are thus almost disjoint unless one is contained in the other. Let F_1 be the union of those cubes in $\mathcal{Q}_1$ which are contained in U, and, inductively, let F_k be the union of those cubes in $\mathcal{Q}_k$ which are contained in U and whose interiors are disjoint from $\bigcup_1^{k-1} F_j$. Then $\bigcup_1^\infty F_k$ is a countable almost disjoint union of cubes contained in U, and we claim that it equals U. Indeed, if $x \in U$ then $\delta = \inf\{\|x - y\|: y \in U\}$ is positive, and if $2^{-k} < \delta$ the cubes $Q \in \mathcal{Q}_k$ which contain x are subsets of U. It therefore follows from the construction of F_k that these cubes are either subsets of F_k or subsets of cubes contained in $\bigcup_1^{k-1} F_j$. In either case, $x \in \bigcup_1^k F_j$. □

(2.47) Theorem. Suppose Ω is an open set in $\mathbf{R}^n$ and $G: \Omega \to \mathbf{R}^n$ is a C^1 diffeomorphism.

(a) If f is a Lebesgue measurable function on $G(\Omega)$, then $f \circ G$ is Lebesgue measurable on Ω. If $f \geq 0$ or $f \in L^1(G(\Omega), m)$, then

$$\int_{G(\Omega)} f(x)\,dx = \int_\Omega f \circ G(x) |\det D_x G|\,dx.$$

(b) If $E \subset \Omega$ and $E \in \mathcal{L}^n$, then $G(E) \in \mathcal{L}^n$ and $m(G(E)) = \int_E |\det D_x G|\,dx$.

Proof. As in the proof of Theorem (2.43) it suffices to consider Borel measurable functions and sets. Since G and G^{-1} are both continuous, there are then no measurability problems.

Let Q be a cube in Ω, say $Q = \{x: \|x - a\| \leq h\}$. By the mean value theorem, $g_i(x) - g_i(a) = \sum_j (x_j - a_j)(\partial g_i/\partial x_j)(y)$ for some y on the line segment joining x and a, so that for $x \in Q$, $\|G(x) - G(a)\| \leq h[\sup_{y \in Q} \|D_y G\|]$. In other words, $G(Q)$ is contained in a cube of side length $\sup_{y \in Q} \|D_y G\|$ times that of Q, so that [by Theorem (2.43)], $m(G(Q)) \leq [\sup_{y \in Q} \|D_y G\|]^n m(Q)$. If $T \in GL(n, \mathbf{R})$, we can apply this formula with G replaced by $T^{-1} \circ G$ together with Theorem (2.43) to obtain

(2.48)

$$m(G(Q)) = |\det T| m\big(T^{-1}(G(Q))\big) \leq |\det T| \big[\sup_{y \in Q} \|T^{-1} D_y G\|\big]^n m(Q).$$

Since $D_y G$ is continuous in y, for any $\varepsilon > 0$ we can choose $\delta > 0$ so that $\|(D_z G)^{-1} D_y G\|^n \leq 1 + \varepsilon$ if $\|y - z\| \leq \delta$, $y, z \in Q$. Let us now subdivide Q into almost disjoint subcubes $Q_1, \ldots, Q_N$ whose side lengths are at most δ and whose centers are $x_1, \ldots, x_N$. Applying (2.48) with Q replaced by Q_j and with

$T = D_{x_j} G$, we obtain

$$\begin{aligned} m(G(Q)) &\leq \sum_1^N m\big(G(Q_j)\big) \\ &\leq \sum_1^N |\det D_{x_j} G| \Big[\sup_{y \in Q_j} \|\big(D_{x_j} G\big)^{-1} D_y G\|\Big]^n m(Q_j) \\ &\leq (1 + \varepsilon) \sum_1^N |\det D_{x_j} G| m(Q_j). \end{aligned}$$

This last sum is a Riemann sum for $\int_Q |\det D_x G|\,dx$. Since $D_x G$ is continuous in x, letting $\varepsilon \to 0$ and $\delta \to 0$ we find that

$$m(G(Q)) \leq \int_Q |\det D_x G|\,dx.$$

We claim that this estimate holds for any Borel set in Ω. Indeed, if $U \subset \Omega$ is open, by Lemma (2.46) we can write $U = \bigcup_1^\infty Q_j$ where the Q_j's are almost disjoint cubes; hence

$$m(G(U)) \leq \sum_1^\infty m\big(G(Q_j)\big) \leq \sum_1^\infty \int_{Q_j} |\det D_x G|\,dx = \int_U |\det D_x G|\,dx.$$

Moreover, if $E \subset \Omega$ is any Borel set of finite measure, by Theorem (2.40) there is a decreasing sequence of open sets $U_j \subset \Omega$ of finite measure such that $E \subset \bigcap_1^\infty U_j$ and $m(\bigcap_1^\infty U_j \setminus E) = 0$. Hence by the dominated convergence theorem,

$$\begin{aligned} m(G(E)) &\leq m\left(G\left(\bigcap_1^\infty U_j\right)\right) = \lim m\big(G(U_j)\big) \leq \lim \int_{U_j} |\det D_x G|\,dx \\ &= \int_E |\det D_x G|\,dx. \end{aligned}$$

Finally, since m is σ-finite, it follows from this that $m(G(E)) \leq \int_E |\det D_x G|\,dx$ for any Borel set $E \subset \Omega$.

If $f = \sum a_j \chi_{A_j}$ is a nonnegative simple function on $G(\Omega)$, we therefore have

$$\begin{aligned} \int_{G(\Omega)} f(x)\,dx &= \sum a_j m(A_j) \\ &\leq \sum a_j \int_{G^{-1}(A_j)} |\det D_x G|\,dx = \int_\Omega f \circ G(x) |\det D_x G|\,dx. \end{aligned}$$

Theorem (2.10) and the monotone convergence theorem then imply that

$$\int_{G(\Omega)} f(x)\,dx \le \int_{\Omega} f\circ G(x)|\det D_x G|\,dx$$

for any nonnegative measurable f. But the same reasoning applies with G replaced by G^{-1} and f replaced by $f\circ G$, so that

$$\int f\circ G(x)|\det D_x G|\,dx \le \int_{G(\Omega)} f\circ G\circ G^{-1}(x)|\det D_{G^{-1}(x)}G|\,|\det D_x G^{-1}|\,dx$$

$$= \int_{G(\Omega)} f(x)\,dx.$$

This establishes (a) for $f \ge 0$, and the case $f \in L^1$ follows immediately. Since (b) is just the special case of (a) where $f = \chi_{G(E)}$, the proof is complete. □

The most important nonlinear coordinate systems in $\mathbf{R}^2$ and $\mathbf{R}^3$ are polar coordinates ($x = r\cos\theta$, $y = r\sin\theta$) and spherical coordinates ($x = r\sin\varphi\cos\theta$, $y = r\sin\varphi\sin\theta$, $z = r\cos\varphi$). The preceding theorem, applied to these coordinates, yields the familiar formulas (loosely stated) $dx\,dy = r\,dr\,d\theta$, $dx\,dy\,dz = r^2\sin\varphi\,dr\,d\theta\,d\varphi$. Similar coordinate systems exist in higher dimensions, but they become increasingly complicated as the dimension increases. For most purposes, however, it is sufficient to know that Lebesgue measure is effectively the product of the measure $r^{n-1}\,dr$ on $(0,\infty)$ and a certain measure on the unit sphere ($d\theta$ for $n = 2$, $\sin\varphi\,d\theta\,d\varphi$ for $n = 3$). We shall now construct this measure.

Let $S^{n-1} = \{x \in \mathbf{R}^n\colon |x| = 1\}$. If $x \in \mathbf{R}^n \setminus \{0\}$, we shall employ the notation $r = |x|$, $x' = x/|x|$. The map $\Phi(x) = (r, x')$ is then a continuous bijection from $\mathbf{R}^n \setminus \{0\}$ to $(0,\infty)\times S^{n-1}$ whose (continuous) inverse is $\Phi^{-1}(r, x') = rx'$. We denote by m_* the Borel measure on $(0,\infty)\times S^{n-1}$ induced by Φ from Lebesgue measure, that is, $m_*(E) = m(\Phi^{-1}(E))$. Moreover, we define the measure $\rho = \rho_n$ on $(0,\infty)$ by $\rho(E) = \int_E r^{n-1}\,dr$.

(2.49) Theorem. There is a unique Borel measure $\sigma = \sigma_{n-1}$ on S^{n-1} such that $m_* = \rho\times\sigma$. If f is Borel measurable on $\mathbf{R}^n$ and $f \ge 0$ or $f \in L^1(m)$, then

$$\int_{\mathbf{R}^n} f(x)\,dx = \int_0^\infty \int_{S^{n-1}} f(rx')r^{n-1}\,d\sigma(x')\,dr. \tag{2.50}$$

Proof. Equation (2.50), when f is a characteristic function, is merely a restatement of the equation $m_* = \rho\times\sigma$, and it follows for general f by the usual linearity and approximation arguments; hence we need only prove the

first assertion. If E is a Borel set in S^{n-1}, for $a > 0$ let

$$E_a = \Phi^{-1}((0,a]\times E) = \{rx'\colon 0 < r \le a, x \in E\}.$$

If (2.50) is to hold when $f = \chi_{E_1}$ we must have

$$m(E_1) = \int_0^1\int_E r^{n-1}\,d\sigma(x')\,dr = \sigma(E)\int_0^1 r^{n-1}\,dr = n^{-1}\sigma(E).$$

We therefore *define* $\sigma(E) = nm(E_1)$. Since the map $E \to E_1$ takes Borel sets to Borel sets and commutes with unions, intersections, and complements, it is clear that σ is a Borel measure on S^{n-1}. Also, since E_a is the image of E_1 under the map $x \to ax$, it follows from Theorem (2.43) that $m(E_a) = a^n m(E_1)$, and hence, if $0 < a < b$,

$$m_*((a,b]\times E) = m(E_b \setminus E_a) = n^{-1}(b^n - a^n)\sigma(E) = \sigma(E)\int_a^b r^{n-1}\,dr$$

$$= \rho\times\sigma((a,b]\times E).$$

Fix $E \in \mathscr{B}_{S^{n-1}}$ and let $\mathscr{A}_E$ be the collection of finite disjoint unions of sets of the form $(a,b]\times E$. By Proposition (1.7), $\mathscr{A}_E$ is an algebra on $(0,\infty)\times E$ which generates the σ-algebra $\mathscr{M}_E = \{A\times E\colon A \in \mathscr{B}_{(0,\infty)}\}$. By the preceding calculation we have $m_* = \rho\times\sigma$ on $\mathscr{A}_E$, and hence by the uniqueness assertion of Theorem (1.14), $m_* = \rho\times\sigma$ on $\mathscr{M}_E$. But $\bigcup\{\mathscr{M}_E\colon E \in \mathscr{B}_{S^{n-1}}\}$ is precisely the set of Borel rectangles on $(0,\infty)\times S^{n-1}$, so another application of the uniqueness theorem shows that $m_* = \rho\times\sigma$ on all Borel sets. □

Remark. Of course, (2.50) can be extended to Lebesgue measurable functions by considering the completion of the measure σ. Details are left to the reader.

(2.51) Corollary. If f is a measurable function on $\mathbf{R}^n$, nonnegative or integrable, such that $f(x) = g(|x|)$ for some function g on $(0,\infty)$, then

$$\int f(x)\,dx = \sigma(S^{n-1})\int_0^\infty g(r)r^{n-1}\,dr.$$

(2.52) Corollary. Given $a > 0$, let $B = \{x \in \mathbf{R}^n\colon |x| < a\}$, and let f be a measurable function on $\mathbf{R}^n$.

(a) If $|f(x)| \le C|x|^{-\alpha}$ on B for some $C > 0$ and $\alpha < n$, then $f \in L^1(B)$. If $|f(x)| \ge C|x|^{-n}$ on B, then $f \notin L^1(B)$.

(b) If $|f(x)| \le C|x|^{-\alpha}$ on B^c for some $C > 0$ and $\alpha > n$, then $f \in L^1(B^c)$. If $|f(x)| \ge C|x|^{-n}$ on B^c, then $f \notin L^1(B^c)$.

Proof. Apply Corollary (2.51) to $|x|^{-\alpha}\chi_B$ and $|x|^{-\alpha}\chi_{B^c}$. □

We shall compute $\sigma(S^{n-1})$ below. Of course, we know that $\sigma(S^1) = 2\pi$ (this is just the definition of 2π as the ratio of the circumference of a circle to its radius). Armed with this fact, we can compute a very important integral.

(2.53) Proposition. If $a > 0$,

$$\int_{\mathbf{R}^n} \exp(-a|x|^2)\,dx = \left(\frac{\pi}{a}\right)^{n/2}.$$

Proof. Let $I_n = \int_{\mathbf{R}^n} \exp(-a|x|^2)\,dx$. For $n = 2$, by Corollary (2.51) we have

$$I_2 = 2\pi \int_0^\infty re^{-ar^2}\,dr = -\left(\frac{\pi}{a}\right)e^{-ar^2}\Big|_0^\infty = \frac{\pi}{a}.$$

Since $\exp(-a|x|^2) = \prod_1^n \exp(-ax_j^2)$, Tonelli's theorem implies that $I_n = (I_1)^n$. In particular, $I_1 = (I_2)^{1/2}$, so $I_n = (I_2)^{n/2} = (\pi/a)^{n/2}$. □

Once we know this, the device used in the proof can be turned around to compute $\sigma(S^{n-1})$ for all n. We recall that the gamma function is defined by

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt \quad (x > 0).$$

We have $\Gamma(1) = 1$ and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ [make the substitution $t = s^2$ and use Proposition (2.53)]. Moreover, integration by parts shows that $\Gamma(x+1) = x\Gamma(x)$. It follows that for $n \in \mathbf{N}$,

$$(2.54)\quad \Gamma(n) = (n-1)!, \qquad \Gamma(n+\tfrac{1}{2}) = (n-\tfrac{1}{2})(n-\tfrac{3}{2})\cdots(\tfrac{1}{2})\sqrt{\pi}.$$

(2.55) Proposition. $\sigma(S^{n-1}) = 2\pi^{n/2}/\Gamma(n/2)$.

Proof. By Corollary (2.51), Proposition (2.53), and the substitution $s = r^2$,

$$\pi^{n/2} = \int_{\mathbf{R}^n} e^{-|x|^2}\,dx = \sigma(S^{n-1})\int_0^\infty r^{n-1}e^{-r^2}\,dr$$

$$= \frac{\sigma(S^{n-1})}{2}\int_0^\infty s^{(n-2)/2}e^{-s}\,ds = \frac{\sigma(S^{n-1})}{2}\Gamma\left(\frac{n}{2}\right).$$ □

(2.56) Corollary. If $B^n = \{x \in \mathbf{R}^n\colon |x| < 1\}$, then

$$m(B^n) = \pi^{n/2}/\Gamma((n/2)+1).$$

Proof. $m(B^n) = \sigma(S^{n-1})\int_0^1 r^{n-1}\,dr = \sigma(S^{n-1})/n = 2\pi^{n/2}/n\Gamma(n/2) = \pi^{n/2}/\Gamma((n/2)+1)$. □

It is worth noting that, by (2.54), $\sigma(S^{n-1})$ and $m(B^n)$ are always rational multiples of integral powers of π, and the power of π increases by 1 when n increases by 2.

EXERCISES

53. Fill in the details of the proof of Theorem (2.41).

54. Theorem (2.43) remains valid for noninvertible linear maps T. [The statement of part (a) needs to be modified slightly.]

55. The measure σ on S^{n-1} is rotation invariant.

56. The technique used to prove Proposition (2.55) can also be used to integrate any polynomial over S^{n-1}. More precisely, let $f(x) = \prod_1^n x_j^{\alpha_j}$ ($\alpha_j = 0, 1, 2, \dots$). Then $\int f\,d\sigma = 0$ if any α_j is odd, and if all α_j's are even,

$$\int f\,d\sigma = \frac{2\prod_1^n \Gamma(\beta_j)}{\Gamma(\beta_1 + \cdots + \beta_n)} \quad \text{where } \beta_j = \frac{\alpha_j + 1}{2}.$$

57. Let $E = [0,1] \times [0,1]$. Investigate the existence and equality of $\int_E f\,dm^2$, $\int_0^1\int_0^1 f(x,y)\,dx\,dy$, and $\int_0^1\int_0^1 f(x,y)\,dy\,dx$ for the following f:
(a) $f(x,y) = (x^2 - y^2)(x^2 + y^2)^{-2}$.
(b) $f(x,y) = (1 - xy)^{-a}$ $(a > 0)$.
(c) $f(x,y) = (x - \frac{1}{2})^{-3}$ if $0 < y < |x - \frac{1}{2}|$, $f(x,y) = 0$ otherwise.

58. If f is Lebesgue integrable on $(0,a)$ and $g(x) = \int_x^a t^{-1}f(t)\,dt$, then g is integrable on $(0,a)$, and $\int_0^a g(x)\,dx = \int_0^a f(x)\,dx$.

59. By integrating $e^{-sxy}\sin x$ with respect to x and y, show that $\int_0^\infty e^{-sx}x^{-1}\sin x\,dx = \arctan(1/s)$ for $s > 0$. [It may be useful to recall that $\tan((\pi/2) - \theta) = 1/\tan\theta$. Cf. Exercise 31d.]

60. By integrating $e^{-sx}\sin 2xy$ with respect to x and y, show that $\int_0^\infty e^{-sx}x^{-1}\sin^2 x\,dx = \frac{1}{4}\log(1 + (4/s^2))$ for $s > 0$.

61. (a) Show that $\int_0^\infty |x^{-1}\sin x|\,dx = \infty$.
(b) By integrating $e^{-xy}\sin x$ with respect to x and y, show that $\lim_{b\to\infty}\int_0^b x^{-1}\sin x\,dx = \pi/2$. [In view of part (a), some care is needed in passing to the limit as $b \to \infty$.]

62. Show that $\Gamma(x)\Gamma(y)/\Gamma(x+y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$ for $x, y > 0$. [Write $\Gamma(x)\Gamma(y)$ as a double integral and use the argument of the exponential as a new variable of integration.]

63. If f is continuous on $[0, \infty)$, for $\alpha > 0$ and $x \geq 0$ let

$$I_\alpha f(x) = \Gamma(\alpha)^{-1} \int_0^x (x - t)^{\alpha - 1} f(t)\, dt.$$

$I_\alpha f$ is called the αth **fractional integral** of f.

(a) $I_{\alpha+\beta} f = I_\alpha(I_\beta f)$ for all $\alpha, \beta > 0$. (Use Exercise 62.)

(b) If $n \in \mathbf{N}$, $I_n f$ is an nth order antiderivative of f.

7. NOTES AND REFERENCES

The history of modern measure and integration theory can fairly be said to have begun with the publication of Lebesgue's thesis [62] in 1902, although of course Lebesgue was building on earlier works of other mathematicians, and some of his results were obtained independently by Vitali and W. H. Young. The theory of the Lebesgue integral was extensively developed by a number of mathematicians in the ensuing decade, during which time most of the results in this chapter were first derived. In particular, Lebesgue himself proved the dominated convergence theorem and deduced the monotone convergence theorem from it in the case when the limit function f is integrable; when $\int f = \infty$ the latter theorem is due to B. Levi.

Lebesgue [63] studied more general measures on $\mathbf{R}^n$ (which he called "additive set functions") in connection with the problem of generalizing the notion of indefinite integrals to functions of several variables. Radon [77] then developed the theory of integration with respect to what we now call regular Borel measures on $\mathbf{R}^n$, which in particular yields the Lebesgue–Stieltjes integrals when $n = 1$. Finally, in 1915 Fréchet [36] pointed out that many of Radon's ideas would work in the general setting of sets equipped with σ-algebras. Thus was abstract measure and integration theory born. It continued to develop until, by the early 1950s, it had assumed more or less the form in which we know it today.

For accounts of the prehistory and early history of the Lebesgue integral, see Hawkins [49] and Lebesgue [64]. References concerning the later development of the subject can be found in Saks [93] and Hahn and Rosenthal [40].

We have adopted the point of view of beginning with measures and deriving integrals from them. However, it is also possible to go the other way, a procedure first developed by Daniell [21]. Roughly speaking, one starts with an "elementary integral": a linear functional I defined on a suitable space of functions which satisfies some mild continuity conditions and is positive in the sense that $I(f) \geq 0$ whenever $f \geq 0$ (for example, the Riemann integral on the space of continuous functions on $[a, b]$). The Daniell theory provides an extension of I to a functional $\bar{I}$ defined on a larger class of functions. Under appropriate hypotheses, the collection $\mathcal{M}$ of sets E such that χ_E is in the domain of $\bar{I}$ is then a σ-algebra, the function $\mu(E) = \bar{I}(\chi_E)$ is a measure on $\mathcal{M}$, and $\bar{I}$ is integration with respect to μ. See Royden [87] for a concise account of the Daniell theory and Pfeffer [76] for a comprehensive treatment.

It is sometimes of interest to integrate vector-valued functions. For functions with values in $\mathbf{R}^n$ or $\mathbf{C}^n$, one simply integrates each component. There are also theories of integration for functions with values in certain infinite-dimensional vector spaces; one of these is outlined in Exercises 15–20 in Chapter 5.

Section 1. There is a hierarchy of Borel measurable functions on a metric space which corresponds roughly to the hierarchy of Borel sets (open and closed, F_σ and G_δ, etc.). Namely, let B_0 be the space of all continuous functions, and for each countable ordinal α define B_α recursively as follows. If α has an immediate predecessor β, B_α is the set of all limits of pointwise convergent sequences in B_β; otherwise, $B_\alpha = \bigcup_{\beta < \alpha} B_\beta$. Functions in B_α are said to be of **Baire class** α. For example, if f is everywhere differentiable on $\mathbf{R}$, f' is of Baire class 1.

Exercise 11 is a result from Lebesgue's first published paper. See Rudin [91] for a discussion of it.

Section 4. An interesting discussion of Egoroff's theorem, including some necessary and sufficient conditions for almost uniform convergence, can be found in Bartle [9]. For a simple proof of Lusin's theorem (Exercise 44) which does not depend on Egoroff's theorem, see Feldman [30]. We shall prove a more general form of this theorem in Section 7.2.

Section 5. The original theorems of Fubini and Tonelli pertained to Lebesgue measure in the plane. The theory of abstract product measures was developed independently by several people in the early 1930s; the construction of $\mu \times \nu$ presented here is that of Hahn [39]. It is also possible to define a product measure on the product of an infinite family $\{(X_\alpha, \mathcal{M}_\alpha, \mu_\alpha)\}_{\alpha \in A}$ of measures spaces provided that $\mu_\alpha(X_\alpha) = 1$ for all but finitely many α. We shall present a version of this result in Section 9.4.

Using the axiom of choice but not the continuum hypothesis, Sierpiński [98] has proved the existence of a Lebesgue nonmeasurable subset of $\mathbf{R}^2$ whose intersection with any straight line contains at most two points. This should be compared with Exercise 47 (which is also due to Sierpiński).

Section 6. Our proof of Theorem (2.47) follows J. Schwartz [95]. This theorem can also be proved under slightly weaker hypotheses on the transformation G; see Rudin [90], Theorem 8.26.

The surface measure σ on S^{n-1} can be given explicitly as follows. One parametrizes S^{n-1} by angles $\varphi_1, \ldots, \varphi_{n-2}, \theta$ $(0 \leq \varphi_j \leq \pi, 0 \leq \theta \leq 2\pi)$ which are related to Cartesian coordinates $x_1, \ldots, x_n$ by

$$x_1 = \cos\varphi_1, \quad x_2 = \sin\varphi_1 \cos\varphi_2, \quad x_3 = \sin\varphi_1 \sin\varphi_2 \cos\varphi_3, \ldots,$$

$$x_{n-1} = \sin\varphi_1 \ldots \sin\varphi_{n-2} \cos\theta, \quad x_n = \sin\varphi_1 \ldots \sin\varphi_{n-2} \sin\theta.$$

Then σ is given by

$$d\sigma(\varphi_1, \ldots, \varphi_{n-2}, \theta) = \sin^{n-2}\varphi_1 \sin^{n-3}\varphi_2 \ldots \sin\varphi_{n-2}\, d\varphi_1 \ldots d\varphi_{n-2}\, d\theta.$$

3

DECOMPOSITION AND DIFFERENTIATION OF MEASURES

The principal theme of this chapter is the concept of differentiating a measure ν with respect to another measure μ defined on the same σ-algebra. We do this first on the abstract level, then obtain a more refined result when μ is Lebesgue measure on $\mathbf{R}^n$. When the latter theorem is specialized to the case $n = 1$, it joins with classical real variable theory to produce a version of the fundamental theorem of calculus for Lebesgue integrals.

In developing this program it is useful to generalize the notion of measure so as to allow measures to assume negative or even complex values. There are three reasons for this. First, in applications such "signed measures" can represent things such as electric charge which can be either positive or negative. Second, the differentiation theory proceeds more naturally in the more general setting. Finally, complex measures have a functional-analytic significance which will be explained in Chapter 7.

1. SIGNED MEASURES

Let $(X, \mathcal{M})$ be a measurable space. A **signed measure** on $(X, \mathcal{M})$ is a function $\nu\colon \mathcal{M} \to [-\infty, \infty]$ such that (i) $\nu(\varnothing) = 0$, (ii) ν assumes at most one of the values $\pm\infty$, and (iii) if $\{E_j\}$ is a sequence of disjoint sets in $\mathcal{M}$, then $\nu(\bigcup_1^\infty E_j) = \sum_1^\infty \nu(E_j)$, where the latter series converges absolutely if $\nu(\bigcup_1^\infty E_j)$ is finite. Thus every measure is a signed measure; for emphasis we shall sometimes refer to measures as **positive** measures.

Two examples of signed measures come readily to mind. First, if μ_1, μ_2 are measures on $\mathcal{M}$ and at least one of them is finite, then $\nu = \mu_1 - \mu_2$ is a signed measure. Second, if μ is a measure on $\mathcal{M}$ and $f\colon X \to [-\infty, \infty]$ is a measurable function such that at least one of $\int f^+ d\mu$ and $\int f^- d\mu$ is finite (in

which case we shall call f an **extended μ-integrable** function), then the set function ν defined by $\nu(E) = \int_E f d\mu$ is a signed measure. In fact, we shall see shortly that every signed measure can be represented in either of these two forms.

(3.1) Proposition. Let ν be a signed measure on $(X, \mathcal{M})$. If $\{E_j\}$ is an increasing sequence in $\mathcal{M}$, then $\nu(\bigcup_1^\infty E_j) = \lim_{j\to\infty} \nu(E_j)$. If $\{E_j\}$ is a decreasing sequence in $\mathcal{M}$ and $\nu(E_1)$ is finite, then $\nu(\bigcap_1^\infty E_j) = \lim_{j\to\infty} \nu(E_j)$. □

The proof is essentially the same as for positive measures [Theorem (1.8)]; details are left to the reader (Exercise 1).

If ν is a signed measure on $(X, \mathcal{M})$, a set $E \in \mathcal{M}$ is called **positive** (resp. **negative, null**) for ν if $\nu(F) \geq 0$ [resp. $\nu(F) \leq 0$, $\nu(F) = 0$] for all $F \in \mathcal{M}$ such that $F \subset E$. [Thus, in the example $\nu(E) = \int_E f d\mu$ mentioned above, E is positive, negative, or null precisely when $f \geq 0$, $f \leq 0$, or $f = 0$ μ-a.e. on E.]

(3.2) Lemma. If P_j is a positive set for ν for all $j \in \mathbf{N}$, then $\bigcup_1^\infty P_j$ is a positive set for ν.

Proof. Let $Q_n = P_n \setminus \bigcup_1^{n-1} P_j$. Then $Q_n \subset P_n$, so Q_n is positive. Hence if $E \subset \bigcup_1^\infty P_j$, then $\nu(E) = \sum_1^\infty \nu(E \cap Q_n) \geq 0$, as desired. □

(3.3) The Hahn Decomposition Theorem. If ν is a signed measure on $(X, \mathcal{M})$, there exist a positive set P and a negative set N for ν such that $P \cup N = X$ and $P \cap N = \varnothing$. If P', N' is another such pair, then $P \Delta P'$ $(= N \Delta N')$ is null for ν.

Proof. Without loss of generality, we assume that ν does not assume the value $+\infty$. (Otherwise, consider $-\nu$.) We first claim that if $\nu(A) > -\infty$, for every $\varepsilon > 0$ there exists $B \subset A$ such that $\nu(B) \geq \nu(A)$ and $\nu(E) > -\varepsilon$ for all $E \subset B$. If not, there exists $E_1 \subset A$ with $\nu(E_1) \leq -\varepsilon$. Since $\nu(A \setminus E_1) = \nu(A) - \nu(E_1) \geq \nu(A)$, there exists $E_2 \subset A \setminus E_1$ with $\nu(E_2) \leq -\varepsilon$. Continuing inductively, we obtain disjoint sets $E_j \subset A$ $(j \in \mathbf{N})$ with $\nu(E_j) \leq -\varepsilon$. But then if $E = \bigcup_1^\infty E_j$ we have $\nu(A \setminus E) = \nu(A) - \sum_1^\infty \nu(E_j) = \infty$, contrary to assumption.

Next we claim that if $\nu(A) > -\infty$ there is a positive set $P \subset A$ with $\nu(P) \geq \nu(A)$. Indeed, let $A_1 = A$ and define A_n inductively as follows. Having picked $A_1, \ldots, A_{n-1}$, by the preceding remarks we can find $A_n \subset A_{n-1}$ such that $\nu(A_n) \geq \nu(A_{n-1})$ and $\nu(E) \geq -1/n$ for all $E \subset A_n$. Let $P = \bigcap_1^\infty A_n$. Then P is positive by construction, and by Proposition (3.1), $\nu(P) = \lim \nu(A_n) \geq \nu(A)$.

Now let $s = \sup\{\nu(A)\colon A \in \mathcal{M}\}$. There is a sequence $\{P_n\}$ such that $\nu(P_n) \to s$, and by the claim established above we can assume that P_n is positive for each n. Let $P = \bigcup_1^\infty P_n$. Then $\nu(P) = s$ and P is positive, by

proposition (3.1) and Lemma (3.2). Moreover, $N = P^c$ is negative, for if $E \subset N$ and $\nu(E) > 0$, we would have $\nu(P \cup E) = \nu(P) + \nu(E) > s$, which is impossible.

Finally, if P', N' are another such pair we have $P \setminus P' \subset P$ and $P \setminus P' \subset N'$, so that $P \setminus P'$ is both positive and negative, hence null. Likewise for $P' \setminus P$. □

The decomposition $X = P \cup N$ of X as the disjoint union of a positive set and a negative set is called a **Hahn decomposition** for ν. It is usually not unique, but it leads to a canonical representation of ν as the difference of two positive measures. To state this result we need to introduce a new concept: we say that two signed measures μ and ν on $(X, \mathcal{M})$ are **mutually singular**, or that ν is **singular with respect to** μ, or vice versa, if there exist $E, F \in \mathcal{M}$ such that $E \cap F = \varnothing$, $E \cup F = X$, E is null for μ, and F is null for ν. Informally speaking, mutual singularity means that μ and ν "live on disjoint sets." We express this relationship symbolically by writing $\mu \perp \nu$.

(3.4) Theorem. If ν is a signed measure, there exist unique positive measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$.

Proof. Let $X = P \cup N$ be a Hahn decomposition for ν, and set $\nu^+(E) = \nu(E \cap P)$, $\nu^-(E) = -\nu(E \cap N)$. Then clearly $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$. If also $\nu = \mu^+ - \mu^-$ where $\mu^+ \perp \mu^-$, let $E, F \in \mathcal{M}$ be such that $E \cap F = \varnothing$, $E \cup F = X$, and $\mu^+(F) = \mu^-(E) = 0$. Then $X = E \cup F$ is a Hahn decomposition for ν, so $P \Delta E$ is ν-null. Therefore, for any $A \in \mathcal{M}$, $\mu^+(A) = \mu^+(A \cap E) = \nu(A \cap E) = \nu(A \cap P) = \nu^+(A)$, and likewise $\nu^- = \mu^-$. □

The measures ν^+ and ν^- are called the **positive** and **negative variations** of ν, and $\nu = \nu^+ - \nu^-$ is called the **Jordan decomposition** of ν, by analogy with the representation of a function of bounded variation on $\mathbf{R}$ as the difference of two increasing functions (see Section 5). Furthermore, we define the **total variation** of ν to be the measure $|\nu|$ defined by

$$|\nu| = \nu^+ + \nu^-.$$

It is easily verified that $E \in \mathcal{M}$ is ν-null iff $|\nu|(E) = 0$, and $\nu \perp \mu$ iff $|\nu| \perp \mu$ iff $\nu^+ \perp \mu$ and $\nu^- \perp \mu$ (Exercise 3).

We observe that if ν omits the value ∞ then $\nu^+(X) = \nu(P) < \infty$, so that ν^+ is a finite measure and ν is bounded above by $\nu^+(X)$; similarly if ν omits the value $-\infty$. In particular, if the range of ν is contained in $\mathbf{R}$, then ν is bounded. We observe also that ν is of the form $\nu(E) = \int_E f d\mu$ where $\mu = |\nu|$ and $f = \chi_P - \chi_N$, $X = P \cup N$ being a Hahn decomposition for ν.

Integration with respect to a signed measure ν is defined in the obvious way: we set

$$L^1(\nu) = L^1(\nu^+) \cap L^1(\nu^-),$$

$$\int f d\nu = \int f d\nu^+ - \int f d\nu^- \quad \text{for } f \in L^1(\nu).$$

One more piece of terminology: a signed measure ν is called **finite** (resp. **σ-finite**) if $|\nu|$ is finite (resp. σ-finite).

EXERCISES

1. Prove Proposition (3.1).
2. If ν is a signed measure, E is ν-null iff $|\nu|(E) = 0$.
3. If ν and μ are signed measures, $\nu \perp \mu$ iff $|\nu| \perp \mu$ iff $\nu^+ \perp \mu$ and $\nu^- \perp \mu$.
4. Let ν be a signed measure on $(X, \mathcal{M})$.
   (a) $L^1(\nu) = L^1(|\nu|)$.
   (b) If $f \in L^1(\nu)$, $|\int f d\nu| \le \int |f| d\nu$.
   (c) If $E \in \mathcal{M}$, $|\nu|(E) = \sup\{|\int_E f d\nu| : |f| \le 1\}$.
5. (a) If ν is a signed measure and λ, μ are positive measures such that $\nu = \lambda - \mu$, then $\lambda \ge \nu^+$ and $\mu \ge \nu^-$.
   (b) If ν_1, ν_2 are signed measures, then $|\nu_1 + \nu_2| \le |\nu_1| + |\nu_2|$.
6. If $\nu(E) = \int_E f d\mu$ where μ is a positive measure, describe the Hahn decompositions of ν and the positive, negative, and total variations of ν.
7. Suppose ν is a signed measure on $(X, \mathcal{M})$ and $E \in \mathcal{M}$.
   (a) $\nu^+(E) = \sup\{\nu(F) : F \in \mathcal{M}, F \subset E\}$, and $\nu^-(E) = -\inf\{\nu(F) : F \in \mathcal{M}, F \subset E\}$.
   (b) $|\nu|(E) = \sup\{\sum_1^n |\nu(E_j)| : E_1, \ldots, E_n$ are disjoint and $\cup_1^n E_j = E\}$.

2. THE LEBESGUE–RADON–NIKODYM THEOREM

Suppose ν is a signed measure and μ is a positive measure on $(X, \mathcal{M})$. We say that ν is **absolutely continuous** with respect to μ and write $\nu \ll \mu$ if $\nu(E) = 0$ for every $E \in \mathcal{M}$ such that $\mu(E) = 0$. It is easily verified that $\nu \ll \mu$ iff $|\nu| \ll \mu$ iff $\nu^+ \ll \mu$ and $\nu^- \ll \mu$ (Exercise 8). Absolute continuity is in a sense the antithesis of mutual singularity: if $\nu \perp \mu$ and $\nu \ll \mu$ then $\nu = 0$, for if E and F are disjoint sets such that $X = E \cup F$ and $\mu(E) = |\nu|(F) = 0$, then the fact that $\nu \ll \mu$ implies that $|\nu|(E) = 0$, whence $|\nu| = 0$ and $\nu = 0$. One can extend the notion of absolute continuity to the case where μ is a signed

measure: namely, $\nu \ll \mu$ iff $\nu \ll |\mu|$. We shall not need this more general definition, however.

The name "absolute continuity" is derived from real variable theory: see Section 5. For finite signed measures it is equivalent to another condition which is obviously a form of continuity:

(3.5) Theorem. Let ν be a finite signed measure and μ a positive measure on $(X, \mathcal{M})$. Then $\nu \ll \mu$ iff for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|\nu(E)| < \varepsilon$ whenever $\mu(E) < \delta$.

Proof. Since $\nu \ll \mu$ iff $|\nu| \ll \mu$ and $|\nu(E)| \le |\nu|(E)$, it suffices to assume that $\nu = |\nu|$ is positive. Clearly the ε-δ condition implies that $\nu \ll \mu$. On the other hand, suppose there exists $\varepsilon > 0$ such that for all $n \in \mathbf{N}$ we can find $E_n \in \mathcal{M}$ with $\mu(E_n) < 2^{-n}$ and $\nu(E_n) \ge \varepsilon$. Let $F_k = \bigcup_k^\infty E_n$ and $F = \cap_1^\infty F_k$. Then $\mu(F_k) < 2^{1-k}$, so $\mu(F) = 0$; but $\nu(F_k) \ge \varepsilon$ and hence, since ν is finite, $\nu(F) = \lim \nu(F_n) \ge \varepsilon$. Thus it is false that $\nu \ll \mu$. □

If μ is a measure and f is an extended μ-integrable function, the signed measure ν defined by $\nu(E) = \int_E f\,d\mu$ is clearly absolutely continuous with respect to μ; it is finite iff $f \in L^1(\mu)$. For any complex-valued $f \in L^1(\mu)$, the preceding theorem can be applied to $\operatorname{Re} f$ and $\operatorname{Im} f$, and we obtain:

(3.6) Corollary. If $f \in L^1(\mu)$, for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|\int_E f\,d\mu| < \varepsilon$ whenever $\mu(E) < \delta$.

We shall use the following notation to express the relationship $\nu(E) = \int_E f\,d\mu$:

$$d\nu = f\,d\mu.$$

Sometimes, by a slight abuse of language, we shall refer to "the signed measure $f\,d\mu$."

We now come to the main results of this section.

(3.7) Lemma. Suppose ν and μ are finite measures on $(X, \mathcal{M})$. Either $\nu \perp \mu$ or there exist $\varepsilon > 0$ and $E \in \mathcal{M}$ such that $\mu(E) > 0$ and $\nu \ge \varepsilon\mu$ on E (i.e., E is a positive set for $\nu - \varepsilon\mu$).

Proof. Let $X = P_n \cup N_n$ be a Hahn decomposition for $\nu - n^{-1}\mu$, and let $P = \cup_1^\infty P_n$ and $N = \cap_1^\infty N_n = P^c$. Then N is a negative set for $\nu - n^{-1}\mu$ for all n, i.e., $0 \le \nu(N) \le n^{-1}\mu(N)$ for all n, so $\nu(N) = 0$. If $\mu(P) = 0$, then $\nu \perp \mu$. If $\mu(P) > 0$, then $\mu(P_n) > 0$ for some n, and P_n is a positive set for $\nu - n^{-1}\mu$. □

(3.8) The Lebesgue–Radon–Nikodym Theorem. Let ν be a σ-finite signed measure and μ a σ-finite positive measure on $(X, \mathcal{M})$. There exist unique

σ-finite signed measures λ, ρ on $(X, \mathcal{M})$ such that $\lambda \perp \mu$, $\rho \ll \mu$, and $\nu = \lambda + \rho$. Moreover, there is an extended μ-integrable function $f\colon X \to \mathbf{R}$, unique modulo alterations on μ-null sets, such that $d\rho = f\,d\mu$.

Proof. First suppose that ν and μ are finite positive measures. Let

$$\mathcal{F} = \left\{ f\colon X \to [0, \infty]\colon \int_E f\,d\mu \le \nu(E) \text{ for all } E \in \mathcal{M} \right\}.$$

We observe that if $f, g \in \mathcal{F}$, then $h = \max(f, g) \in \mathcal{F}$, for if $A = \{x\colon f(x) > g(x)\}$, for any $E \in \mathcal{M}$ we have

$$\int_E h\,d\mu = \int_{E\cap A} f\,d\mu + \int_{E\setminus A} g\,d\mu \le \nu(E \cap A) + \nu(E \setminus A) = \nu(E).$$

Let $a = \sup\{\int_X f\,d\mu\colon f \in \mathcal{F}\}$, noting that $a \le \nu(X) < \infty$, and choose a sequence $\{f_n\} \subset \mathcal{F}$ such that $\int_X f_n\,d\mu \to a$. Let $g_n = \max(f_1, \dots, f_n)$ and $f = \sup_n f_n$. Then $g_n \in \mathcal{F}$, g_n increases pointwise to f, and $\int g_n\,d\mu \ge \int f_n\,d\mu$. It follows that $\lim \int g_n\,d\mu = a$ and hence, by the monotone convergence theorem, that $f \in \mathcal{F}$ and $\int f\,d\mu = a$. (In particular, $f < \infty$ a.e., so we may take f to be real-valued everywhere.) We claim that the measure $d\lambda = d\nu - f\,d\mu$ (which is positive since $f \in \mathcal{F}$) is singular with respect to μ. If not, by Lemma (3.7) there exist $E \in \mathcal{M}$ and $\varepsilon > 0$ such that $\mu(E) > 0$ and $\lambda \ge \varepsilon\mu$ on E. But then $\varepsilon\chi_E\,d\mu \le d\lambda = d\nu - f\,d\mu$, that is, $(f + \varepsilon\chi_E)d\mu \le d\nu$, so $f + \varepsilon\chi_E \in \mathcal{F}$ and $\int(f + \varepsilon\chi_E)d\mu = a + \varepsilon\mu(E) > a$, contradicting the definition of a. Finally, if also $d\nu = d\lambda' + f'd\mu$ we have $d\lambda - d\lambda' = (f - f')d\mu$. But $\lambda - \lambda' \perp \mu$ (cf. Exercise 9), while $(f - f')d\mu \ll \mu$; hence $d\lambda - d\lambda' = (f - f')d\mu = 0$, so that $\lambda = \lambda'$ and [by Proposition (2.23)] $f = f'$ μ-a.e. Thus we are done in the case where μ and ν are finite measures.

Next, suppose μ and ν are σ-finite measures. Then X is a countable disjoint union of μ-finite sets and a countable disjoint union of ν-finite sets; by considering intersections of these we obtain a disjoint sequence $\{A_j\} \subset \mathcal{M}$ such that $\mu(A_j) < \infty$ and $\nu(A_j) < \infty$ for all j and $X = \cup_1^\infty A_j$. Define $\mu_j(E) = \mu(E \cap A_j)$ and $\nu_j(E) = \nu(E \cap A_j)$: then by the reasoning above we have $d\nu_j = d\lambda_j + f_j\,d\mu_j$ where $\lambda_j(A_j^c) = 0$ and $f_j = 0$ on A_j^c. Let $\lambda = \Sigma_1^\infty \lambda_j$, $f = \Sigma_1^\infty f_j$. Then $\lambda \perp \mu$ (cf. Exercise 9) and $d\nu = d\lambda + f\,d\mu$; uniqueness follows as before.

Finally, if ν is a signed measure we apply the preceding argument to ν^+ and ν^- and subtract the results. □

The decomposition $\nu = \lambda + \rho$ where $\lambda \perp \mu$ and $\rho \ll \mu$ is called the **Lebesgue decomposition** of ν with respect to μ. In the case where $\nu \ll \mu$, Theorem (3.8) says that $d\nu = f\,d\mu$ for some f. This result is usually known as the **Radon–Nikodym theorem**, and f is called the **Radon–Nikodym derivative** of ν with respect to μ. We denote it by $d\nu/d\mu$. (Strictly speaking, $d\nu/d\mu$

should be construed as the class of functions equal to f μ-a.e.) The formulas suggested by the differential notation $d\nu/d\mu$ are generally correct. For example, it is obvious that $d(\nu_1 + \nu_2)/d\mu = (d\nu_1/d\mu) + (d\nu_2/d\mu)$, and we have the chain rule:

(3.9) Proposition. Suppose ν is a σ-finite signed measure and μ, λ are σ-finite measures on $(X, \mathcal{M})$ such that $\nu \ll \mu$ and $\mu \ll \lambda$.

(a) If $g \in L^1(\nu)$, then $g(d\nu/d\mu) \in L^1(\mu)$ and $\int g\, d\nu = g(d\nu/d\mu)\, d\mu$.

(b) $\nu \ll \lambda$, and $(d\nu/d\lambda) = (d\nu/d\mu)(d\mu/d\lambda)$ λ-a.e.

Proof. By considering ν^+ and ν^- separately we may assume that $\nu \geq 0$. The equation $\int g\, d\nu = \int g(d\nu/d\mu)\, d\mu$ is true when $g = \chi_E$ by definition of $d\nu/d\mu$. It is therefore true for simple functions by linearity, then for nonnegative measurable functions by the monotone convergence theorem, and finally for functions in $L^1(\nu)$ by linearity again. Replacing ν, μ by μ, λ and setting $g = \chi_E\, d\nu/d\mu$ we obtain

$$\nu(E) = \int_E \frac{d\nu}{d\mu}\, d\mu = \int_E \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}\, d\lambda$$

for all $E \in \mathcal{M}$, whence $(d\nu/d\lambda) = (d\nu/d\mu)(d\mu/d\lambda)\lambda$-a.e. □

(3.10) Corollary. If $\mu \ll \lambda$ and $\lambda \ll \mu$, then $d\lambda/d\mu = (d\mu/d\lambda)^{-1}$ a.e. (with respect to either λ or μ).

Nonexample: Let μ be Lebesgue measure and ν the point mass at 0 on $(\mathbf{R}, \mathcal{B}_{\mathbf{R}})$. Clearly $\nu \perp \mu$. The nonexistent Radon–Nikodym derivative $d\nu/d\mu$ is popularly known as the Dirac δ-function.

We conclude this section with a simple but important observation:

(3.11) Proposition. If $\mu_1, \ldots, \mu_n$ are measures on $(X, \mathcal{M})$, there is a measure μ such that $\mu_j \ll \mu$ for all j—namely, $\mu = \sum_1^n \mu_j$.

The proof is trivial.

EXERCISES

8. $\nu \ll \mu$ iff $|\nu| \ll \mu$ iff $\nu^+ \ll \mu$ and $\nu^- \ll \mu$.

9. If $\nu_j \perp \mu$ for $j \in \mathbf{N}$, then $\sum_1^\infty \nu_j \perp \mu$; if $\nu_j \ll \mu$ for $j \in \mathbf{N}$, then $\sum_1^\infty \nu_j \ll \mu$.

10. Theorem (3.5) may fail when ν is not finite. [Consider $d\nu = dx/x$ and $d\mu = dx$ on $(0, 1)$, or ν = counting measure and $\mu(E) = \sum_{n \in E} 2^{-n}$ on $\mathbf{N}$.]

11. Let μ be a positive measure. A collection of functions $\{f_\alpha\}_{\alpha \in A} \subset L^1(\mu)$ is called **uniformly integrable** if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|\int_E f_\alpha\, d\mu| < \varepsilon$ whenever $\mu(E) < \delta$ and $\alpha \in A$.

(a) Any finite subset of $L^1(\mu)$ is uniformly integrable.

(b) If $\{f_n\}$ is a sequence in $L^1(\mu)$ which converges in the L^1 metric to $f \in L^1(\mu)$, then $\{f_n\}$ is uniformly integrable.

12. Let μ_1, ν_1 be σ-finite measures on $(X_1, \mathcal{M}_1)$ and μ_2, ν_2 σ-finite measures on $(X_2, \mathcal{M}_2)$ such that $\nu_1 \ll \mu_1$ and $\nu_2 \ll \mu_2$. Then $\nu_1 \times \nu_2 \ll \mu_1 \times \mu_2$ and

$$\frac{d(\nu_1 \times \nu_2)}{d(\mu_1 \times \mu_2)}(x_1, x_2) = \frac{d\nu_1}{d\mu_1}(x_1)\frac{d\nu_2}{d\mu_2}(x_2).$$

13. Let $X = [0, 1]$, $\mathcal{M} = \mathcal{B}_{[0,1]}$, m = Lebesgue measure, and μ = counting measure on $\mathcal{M}$.

(a) $m \ll \mu$ but $dm \neq f\, d\mu$ for any f.

(b) μ has no Lebesgue decomposition with respect to m.

14. If ν is an arbitrary signed measure and μ is a σ-finite measure on $(X, \mathcal{M})$, there exists an extended μ-integrable function $f\colon X \to [0, \infty]$ such that $d\nu = f\, d\mu$. Hints:

(a) It suffices to assume that μ is finite and ν is positive.

(b) With these assumptions, there exists $E \in \mathcal{M}$ which is σ-finite for ν such that $\mu(E) \geq \mu(F)$ for all sets F which are σ-finite for ν.

(c) The Radon–Nikodym theorem applies on E. If $F \cap E = \varnothing$, then either $\nu(F) = \mu(F) = 0$ or $\mu(F) > 0$ and $\nu(F) = \infty$.

15. A measure μ on $(X, \mathcal{M})$ is called **decomposable** if there is a family $\mathcal{F} \subset \mathcal{M}$ such that (i) $\mu(F) < \infty$ for all $F \in \mathcal{F}$, (ii) the members of $\mathcal{F}$ are disjoint and their union is X, (iii) if $E \in \mathcal{M}$ and $\mu(E) < \infty$ then $\mu(E) = \sum_{F \in \mathcal{F}} \mu(E \cap F)$, and (iv) if $E \subset X$ and $E \cap F \in \mathcal{M}$ for all $F \in \mathcal{F}$ then $E \in \mathcal{M}$.

(a) Every σ-finite measure is decomposable.

(b) If μ is decomposable and ν is any measure on $(X, \mathcal{M})$ such that $\nu \ll \mu$, there exists a measurable $f\colon X \to [0, \infty]$ such that $\nu(E) = \int_E f\, d\mu$ for any E which is σ-finite for μ. (Use Exercise 14 if ν is not σ-finite.)

(c) Under the conditions of part (b), if $\nu(A) = 0$ whenever $\nu(A \cap F) = 0$ for all $F \in \mathcal{F}$, then $\nu(E) = \int_E f\, d\mu$ for all $E \in \mathcal{M}$.

16. Suppose μ, ν are finite measures on $(X, \mathcal{M})$ with $\nu \ll \mu$, and let $\lambda = \mu + \nu$. If $f = d\nu/d\lambda$, then $0 \leq f < 1$ μ-a.e. and $d\nu/d\mu = f/(1 - f)$.

17. Let $(X, \mathcal{M}, \mu)$ be a measure space, $\mathcal{N}$ a sub-σ-algebra of $\mathcal{M}$, and $\nu = \mu|\mathcal{N}$. If $f \in L^1(\mu)$ there exists $g \in L^1(\nu)$ (thus g is $\mathcal{N}$-measurable) such that $\int_E f\, d\mu = \int_E g\, d\nu$ for all $E \in \mathcal{N}$, and g is unique modulo

alterations on ν-null sets. (In probability theory, g is called the **conditional expectation** of f on $\mathcal{N}$.)

3. COMPLEX MEASURES

A **complex measure** on a measurable space $(X, \mathcal{M})$ is a map $\nu\colon \mathcal{M} \to \mathbf{C}$ such that (i) $\nu(\varnothing) = 0$, and (ii) if $\{E_j\}$ is a sequence of disjoint sets in $\mathcal{M}$, then $\nu(\cup_1^\infty E_j) = \sum_1^\infty \nu(E_j)$, where the series on the right converges absolutely. In particular, infinite values are not allowed, so a positive measure is a complex measure only if it is finite. Example: if μ is a positive measure and $f \in L^1(\mu)$, $f\,d\mu$ is a complex measure. If ν is a complex measure, we shall write ν_r and ν_i for the real and imaginary parts of ν. ν_r and ν_i are thus signed measures which do not assume the values $\pm\infty$, so they are bounded; hence the range of ν is a bounded subset of $\mathbf{C}$.

The notions we have developed for signed measures generalize easily to complex measures. For example, we define $L^1(\nu)$ to be $L^1(\nu_r) \cap L^1(\nu_i)$; if $f \in L^1(\nu)$, $\int f d\nu = \int f d\nu_r + i\int f d\nu_i$. If ν and μ are complex measures, we say that $\nu \perp \mu$ if $\nu_a \perp \mu_b$ for $a, b = r, i$, and if λ is a positive measure, we say that $\nu \ll \lambda$ if $\nu_r \ll \lambda$ and $\nu_i \ll \lambda$. The theorems of Section 2 also generalize: one has merely to apply them to the real and imaginary parts of ν. In particular:

(3.12) The Lebesgue–Radon–Nikodym Theorem. If ν is a complex measure and μ is a σ-finite measure on $(X, \mathcal{M})$, there exist a unique complex measure λ and an $f \in L^1(\mu)$, unique modulo alternations on μ-null sets, such that $\lambda \perp \mu$ and $d\nu = d\lambda + f\,d\mu$.

As before, if $\lambda = 0$ in Theorem (3.12), we write $f = d\nu/d\mu$.

The **total variation** of a complex measure ν is the positive measure $|\nu|$ determined by the property that if $d\nu = f\,d\mu$ where μ is a positive measure, then $d|\nu| = |f|\,d\mu$. To see that this is well defined, we observe first that every ν is of the form $f\,d\mu$ for some finite measure μ and some $f \in L^1(\mu)$: indeed, we can take $\mu = |\nu_r| + |\nu_i|$ and use the Radon–Nikodym theorem to obtain f. Second, if $d\nu = f_1\,d\mu_1 = f_2\,d\mu_2$, let $\lambda = \mu_1 + \mu_2$. Then by Proposition (3.9),

$$f_1 \frac{d\mu_1}{d\lambda}\,d\lambda = d\nu = f_2 \frac{d\mu_2}{d\lambda}\,d\lambda,$$

and since $d\mu_1/d\lambda$ and $d\mu_2/d\lambda$ are nonnegative,

$$|f_1|d\mu_1 = |f_1|\frac{d\mu_1}{d\lambda}\,d\lambda = |f_2|\frac{d\mu_2}{d\lambda}\,d\lambda = |f_2|d\mu_2.$$

Hence the definition of $|\nu|$ is independent of the choice of f and μ. This definition agrees with the previous definition of $|\nu|$ when ν is a signed measure,

for in that case $d\nu = (\chi_P - \chi_N)d|\nu|$ where $X = P \cup N$ is a Hahn decomposition, and $|\chi_P - \chi_N| = 1$.

(3.13) Proposition. Let ν be a complex measure on $(X, \mathcal{M})$.

(a) $\nu \ll |\nu|$ and $d\nu/d|\nu|$ has absolute value one $|\nu|$-a.e.

(b) $|\nu(E)| \le |\nu|(E)$ for all $E \in \mathcal{M}$.

(c) $L^1(\nu) = L^1(|\nu|)$, and if $f \in L^1(\nu)$, $|\int f d\nu| \le \int |f|\,d|\nu|$.

Proof. That $\nu \ll |\nu|$ is obvious from the definition. If $d\nu = f\,d\mu$ and $g = d\nu/d|\nu|$, we have $f\,d\mu = d\nu = g\,d|\nu| = g|f|\,d\mu$. Thus $g|f| = f$ μ-a.e. and hence $|\nu|$-a.e. But clearly $|f| > 0$ $|\nu|$-a.e., whence $|g| = 1$ $|\nu|$-a.e. This proves (a), and (b) and (c) are left to the reader (Exercise 18). □

(3.14) Proposition. If ν_1, ν_2 are complex measures on $(X, \mathcal{M})$, then $|\nu_1 + \nu_2| \le |\nu_1| + |\nu_2|$.

Proof. By Proposition (3.11) we can write $\nu_j = f_j\,d\mu$, with the same μ, for $j = 1, 2$. But then $d|\nu_1 + \nu_2| = |f_1 + f_2|d\mu \le |f_1|d\mu + |f_2|d\mu = d|\nu_1| + d|\nu_2|$. □

EXERCISES

18. Prove Proposition (3.13 b, c).

19. If ν, μ are complex measures and λ is a positive measure, then $\nu \perp \mu$ iff $|\nu| \perp |\mu|$, and $\nu \ll \lambda$ iff $|\nu| \ll \lambda$.

20. If ν is a complex measure on $(X, \mathcal{M})$ and $\nu(X) = |\nu|(X)$, then $\nu = |\nu|$.

21. Let ν be a complex measure on $(X, \mathcal{M})$. If $E \in \mathcal{M}$, let

$$\mu_1(E) = \sup\left\{\sum_1^n |\nu(E_j)|\colon E_1, \dots, E_n \text{ disjoint}, E = \bigcup_1^n E_j, n \in \mathbf{N}\right\},$$

$$\mu_2(E) = \sup\left\{\sum_1^\infty |\nu(E_j)|\colon E_1, E_2, \dots \text{ disjoint}, E = \bigcup_1^\infty E_j\right\},$$

$$\mu_3(E) = \sup\left\{\left|\int_E f d\nu\right|\colon |f| \le 1\right\}.$$

Then $\mu_1 = \mu_2 = \mu_3 = |\nu|$. [First show that $\mu_1 \le \mu_2 \le \mu_3$. To see that $\mu_3 = |\nu|$, let $f = \overline{d\nu/d|\nu|}$ and apply Proposition (3.13). To see that $\mu_3 \le \mu_1$, approximate f by simple functions.]

4. DIFFERENTIATION ON $\mathbf{R}^n$

The Radon–Nikodym theorem provides an abstract notion of the "derivative" of a signed or complex measure ν with respect to a measure μ. In this section we analyze more deeply the special case where $(X, \mathcal{M}) = (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n})$ and $\mu = m$ is Lebesgue measure. Here one can define a *pointwise* derivative of ν with respect to m in the following way. Let $B(r, x)$ be the open ball of radius r about x in $\mathbf{R}^n$; then one can consider the limit

$$F(x) = \lim_{r \to 0} \frac{\nu(B(r,x))}{m(B(r,x))}$$

when it exists. [One can also replace the balls $B(r, x)$ by other sets which, in a suitable sense, shrink to x in a regular way; we shall examine this point later.] If $\nu \ll m$, so that $d\nu = f\,dm$, then $\nu(B(r,x))/m(B(r,x))$ is simply the average value of f on $B(r, x)$, so one would hope that $F = f$ m-a.e. This turns out to be the case provided that $\nu(B(r,x))$ is finite for all r, x. From the point of view of the function f, this may be regarded as a generalization of the fundamental theorem of calculus: the derivative of the indefinite integral of f (namely ν) is f.

For the remainder of this section, terms such as "integrable" or "almost everywhere" refer to Lebesgue measure unless otherwise specified. We begin our analysis with a technical lemma which is of interest in its own right.

(3.15) Lemma. Let $\mathcal{C}$ be any collection of open balls in $\mathbf{R}^n$, and let $U = \bigcup_{B \in \mathcal{C}} B$. If $c < m(U)$, there exist disjoint $B_1, \dots, B_k \in \mathcal{C}$ such that $\sum_1^k m(B_j) > 3^{-n}c$.

Proof. If $c < m(U)$, by Theorem (2.39) there is a compact set $K \subset U$ with $m(K) > c$, and finitely many of the balls in $\mathcal{C}$—say, $A_1, \dots, A_m$—cover K. Let B_1 be the largest of the A_j's (that is, choose B_1 to have maximal radius), let B_2 be the largest of the A_j's which are disjoint from B_1, and so on until the sequence of A_j's is exhausted. If A_i is not one of the B_j's, there exists j such that $A_i \cap B_j \neq \varnothing$ and the radius of A_i is at most that of B_j; hence $A_i \subset B_j^*$, where B_j^* is the ball concentric with B_j whose radius is three times that of B_j. But then $K \subset \bigcup_j B_j^*$, so

$$c < m(K) \leq \sum_j m(B_j^*) = 3^n \sum_j m(B_j). \qquad \square$$

A measurable function $f\colon \mathbf{R}^n \to \mathbf{C}$ is called **locally integrable** (with respect to Lebesgue measure) if $\int_K |f(x)|\,dx < \infty$ for every compact $K \subset \mathbf{R}^n$. We denote the space of a locally integrable functions by L^1_{loc}. If $f \in L^1_{loc}$, $x \in \mathbf{R}^n$,

and $r > 0$, we define $A_r f(x)$ to be the average value of f on $B(r, x)$:

$$A_r f(x) = \frac{1}{m(B(r,x))} \int_{B(r,x)} f(y)\,dy.$$

(3.16) Lemma. If $f \in L^1_{loc}$, $A_r f(x)$ is continuous in r for each x and measurable in x for each r.

Proof. From the results in Section 2.6 we know that $m(B(r,x)) = cr^n$ where $c = m(B(1,0))$ and $m(S(r,x)) = 0$ where $S(r,x) = \{y\colon |y - x| = r\}$. From the latter fact it follows that $\chi_{B(r,x)} \to \chi_{B(R,x)}$ a.e. as $r \to R$, so that $\int_{B(r,x)} f(y)\,dy$ is continuous in r by the dominated convergence theorem. Hence $A_r f(x) = c^{-1}r^{-n}\int_{B(r,x)} f(y)\,dy$ is continuous in r. Also,

$$A_r f(x) = c^{-1}r^{-n}\int_{\mathbf{R}^n} g_r(x,y) f(y)\,dy \quad \text{where } g_r = \chi_{\{(x,y)\colon |x-y|<r\}}.$$

g_r is clearly measurable, so the measurability of $A_r f$ follows from Fubini's theorem. $\square$

Next, if $f \in L^1_{loc}$, we define its **Hardy–Littlewood maximal function** Hf by

$$Hf(x) = \sup_{r>0} A_r|f|(x) = \sup_{r>0} \frac{1}{m(B(r,x))} \int_{B(r,x)} |f(y)|\,dy.$$

In view of lemma (3.16), Hf is measurable, for its definition would be unaffected if we took the supremum only over the countable set $\mathbf{Q} \cap (0, \infty)$.

(3.17) The Maximal Theorem. There is a constant $C > 0$ such that for all $f \in L^1$ and all $\alpha > 0$,

$$m(\{x\colon Hf(x) > \alpha\}) \leq \frac{C}{\alpha} \int |f(x)|\,dx.$$

Proof. Let $E_\alpha = \{x\colon Hf(x) > \alpha\}$. For each $x \in E_\alpha$ we can choose $r_x > 0$ such that $A_{r_x}|f|(x) > \alpha$. The balls $B(r_x, x)$ cover E_α, so by Lemma (3.15), if $c < m(E_\alpha)$ there exist $x_1, \dots, x_k \in E_\alpha$ such that the balls $B_j = B(r_{x_j}, x_j)$ are disjoint and $\sum_1^k m(B_j) > 3^{-n}c$. But then

$$c < 3^n \sum_1^k m(B_j) \leq \frac{3^n}{\alpha} \sum_1^k \int_{B_j} |f(y)|\,dy \leq \frac{3^n}{\alpha} \int_{\mathbf{R}^n} |f(y)|\,dy.$$

Letting $c \to m(E_\alpha)$, we obtain the desired result. $\square$

With this tool in hand, we now present three successively sharper versions of the fundamental differentiation theorem.

(3.18) Theorem. If $f \in L^1_{loc}$, then $\lim_{r\to 0} A_r f(x) = f(x)$ for a.e. $x \in \mathbf{R}^n$.

Proof. It suffices to show that for $N \in \mathbf{N}$, $A_r f(x) \to f(x)$ for a.e. x such that $|x| \le N$. But for $|x| \le N$ and $r \le 1$ the values of $A_r f(x)$ depend only on the values of $f(y)$ for $|y| \le N+1$, so by replacing f with $f\chi_{B(N+1,0)}$ we may assume that $f \in L^1$. Given $\varepsilon > 0$, by Theorem (2.40) we can find a continuous integrable function g such that $\int |g(y) - f(y)|dy < \varepsilon$. Continuity of g implies that for every $x \in \mathbf{R}^n$ and $\delta > 0$ there exists $r > 0$ such that $|g(y) - g(x)| < \delta$ whenever $|y - x| < r$, and hence

$$|A_r g(x) - g(x)| = \frac{1}{m(B(r,x))}\left|\int_{B(r,x)} (g(y) - g(x))\,dy\right| < \delta.$$

Therefore $A_r g(x) \to g(x)$ as $r \to 0$ for every x, so

$$\begin{aligned}\limsup_{r\to 0} |A_r f(x) - f(x)| &= \limsup_{r\to 0} |A_r(f-g)(x) + (A_r g - g)(x) + (g - f)(x)| \\ &\le H(f-g)(x) + |f-g|(x).\end{aligned}$$

Hence if

$$E_\alpha = \{x: \limsup_{r\to 0} |A_r f(x) - f(x)| > \alpha\}, \qquad F_\alpha = \{x: |f-g|(x) > \alpha\},$$

we have

$$E_\alpha \subset F_{\alpha/2} \cup \{x: H(f-g)(x) > \alpha/2\}.$$

But $\alpha m(F_\alpha) \le \int_{F_\alpha} |f(x) - g(x)|\,dx < \varepsilon$, so by the maximal theorem,

$$m(E_\alpha) \le \frac{2\varepsilon}{\alpha} + \frac{2C\varepsilon}{\alpha}.$$

Since ε is arbitrary, $m(E_\alpha) = 0$ for all $\alpha > 0$, and we are done. □

This theorem can be rephrased as follows: if $f \in L^1_{loc}$,

$$\lim_{r\to 0} \frac{1}{m(B(r,x))} \int_{B(r,x)} [f(x) - f(y)]\,dy = 0 \text{ for a.e. } x. \tag{3.19}$$

Actually, something stronger is true: one can insert absolute values into the integral in (3.19) and still get almost everywhere convergence to zero. That is, let us define the **Lebesgue set** L_f of f to be

$$L_f = \left\{x: \lim_{r\to 0} \frac{1}{m(B(r,x))} \int_{B(r,x)} |f(y) - f(x)|\,dy = 0\right\}.$$

(3.20) Theorem. If $f \in L^1_{loc}$, $m((L_f)^c) = 0$.

Proof. For each $c \in \mathbf{C}$ we can apply Theorem (3.18) to $g_c(x) = |f(x) - c|$ to conclude that, except on a Lebesgue null set E_c, we have

$$\lim_{r\to 0} \frac{1}{m(B(r,x))} \int_{B(r,x)} |f(y) - c|\,dy = |f(x) - c|.$$

Let R be a countable dense subset of $\mathbf{C}$, and let $E = \cup_{c\in R} E_c$. Then $m(E) = 0$, and if $x \notin E$, for any $\varepsilon > 0$ we can choose $c \in R$ with $|f(x) - c| < \varepsilon$, so that $|f(y) - f(x)| < |f(y) - c| + \varepsilon$, and it follows that

$$\limsup_{r\to 0} \frac{1}{m(B(r,x))} \int_{B(r,x)} |f(y) - f(x)|\,dy \le |f(x) - c| + \varepsilon < 2\varepsilon.$$

Since ε is arbitrary, we are done. □

Finally, we consider families of sets more general than balls. A family $\{E_r\}_{r>0}$ of Borel subsets of $\mathbf{R}^n$ is said to **shrink nicely to** $x \in \mathbf{R}^n$ if (i) $E_r \subset B(r,x)$ for each r, and (ii) there is a constant $\alpha > 0$, independent of r, such that $m(E_r) \ge \alpha m(B(r,x))$. The sets E_r need not contain x itself. For example, if U is any Borel subset of $B(1,0)$ such that $m(U) > 0$, and $E_r = \{x + ry: y \in U\}$, then $\{E_r\}$ shrinks nicely to x. Here, then is the final version of the differentiation theorem.

(3.21) The Lebesgue Differentiation Theorem. Suppose $f \in L^1_{loc}$. For every x in the Lebesgue set of f—in particular, for almost every x—we have

$$\lim_{r\to 0} \frac{1}{m(E_r)} \int_{E_r} |f(y) - f(x)|\,dy = 0 \text{ and } \lim_{r\to 0} \frac{1}{m(E_r)} \int_{E_r} f(y)\,dy = f(x)$$

for every family $\{E_r\}_{r>0}$ that shrinks nicely to x.

Proof. For some $\alpha > 0$ we have

$$\begin{aligned}\frac{1}{m(E_r)} \int_{E_r} |f(y) - f(x)|\,dy &\le \frac{1}{m(E_r)} \int_{B(r,x)} |f(y) - f(x)|\,dy \\ &\le \frac{1}{\alpha m(B(r,x))} \int_{B(r,x)} |f(y) - f(x)|\,dy.\end{aligned}$$

The first equation therefore follows from Theorem 3.20, and one sees immediately that it implies the second one by writing the latter in the form (3.19). □

We now return to the study of measures on $\mathbf{R}^n$. A Borel measure ν on $\mathbf{R}^n$ will be called **regular** if (i) $\nu(K) < \infty$ for every compact K, and (ii) $\nu(E) =$

$\inf\{\nu(U)$: U open, $E \subset U\}$ for every Borel set E. (Condition (ii) is actually implied by condition (i). For $n = 1$ this follows from Theorems (1.16) and (1.18), and we shall prove it for arbitrary n in Section 7.2. For the time being, we assume (ii) explicitly.) We observe that by (i), every regular measure is σ-finite. A signed Borel measure ν will be called regular if ν^+ and ν^- are regular, and a complex Borel measure ν will be called regular if ν_r and ν_i are regular.

(3.22) Theorem. Let ν be a regular signed or complex Borel measure on $\mathbf{R}^n$, and let $d\nu = d\lambda + f\,dm$ be its Lebesgue–Radon–Nikodym representation. (Thus $\lambda \perp m$ and $f \in L^1_{loc}$.) Then for m-almost every $x \in \mathbf{R}^n$,

$$\lim_{r \to 0} \frac{\nu(E_r)}{m(E_r)} = f(x)$$

for every family $\{E_r\}$ that shrinks nicely to x.

Proof. It is not hard to verify that the measure $f\,dm$ is regular (Exercise 26), and thus so is λ. In view of Theorem (3.21), therefore, it suffices to show that if λ is regular and $\lambda \perp m$, then for m-almost every x, $\lambda(E_r)/m(E_r) \to 0$ as $r \to 0$ when E_r shrinks nicely to x. It also suffices to take $E_r = B(r, x)$ and to assume that λ is positive, since for some $\alpha > 0$ we have

$$\left|\frac{\lambda(E_r)}{m(E_r)}\right| \le \frac{|\lambda|(E_r)}{m(E_r)} \le \frac{|\lambda|(B(r,x))}{m(E_r)} \le \frac{|\lambda|(B(r,x))}{\alpha m(B(r,x))}.$$

Assuming $\lambda \ge 0$, then, let A be a Borel set such that $\lambda(A) = m(A^c) = 0$, and let

$$F_k = \left\{x \in A: \limsup_{r \to 0} \frac{\lambda(B(r,x))}{m(B(r,x))} > \frac{1}{k}\right\}.$$

We shall show that $m(F_k) = 0$ for all k, and this will complete the proof.

The argument is similar to the proof of the maximal theorem. By regularity of λ, given $\varepsilon > 0$ there is an open $U_\varepsilon \supset A$ such that $\lambda(U_\varepsilon) < \varepsilon$. Each $x \in F_k$ is the center of a ball $B_x \subset U_\varepsilon$ such that $\lambda(B_x) > k^{-1}m(B_x)$. By Lemma (3.15), if $V_\varepsilon = \bigcup_{x \in F_k} B_x$ and $c < m(V_\varepsilon)$ there exist $x_1, \dots, x_J$ such that $B_{x_1}, \dots, B_{x_J}$ are disjoint and

$$c < 3^n \sum_1^J m(B_{x_j}) \le 3^n k \sum_1^J \lambda(B_{x_j}) \le 3^n k \lambda(V_\varepsilon) \le 3^n k \lambda(U_\varepsilon) \le 3^n k \varepsilon.$$

We conclude that $m(V_\varepsilon) \le 3^n k\varepsilon$, and since $F_k \subset V_\varepsilon$ and ε is arbitrary, $m(F_k) = 0$. □

EXERCISES

22. If $f \in L^1$, $f \ne 0$, there exist $C, R > 0$ such that $Hf(x) \ge C|x|^{-n}$ for $|x| > R$. Hence $m(\{x;\ Hf(x) > \alpha\}) \ge C'/\alpha$ when α is small, so the estimate in the maximal theorem is essentially sharp.

23. A useful variant of the Hardy–Littlewood maximal function is

$$H^*f(x) = \sup\left\{m(B)^{-1}\int_B |f(y)|\,dy:\ B \text{ is a ball and } x \in B\right\}.$$

Show that $Hf \le H^*f \le 2^n Hf$.

24. If $f \in L^1_{loc}$ and f is continuous at x, then x is in the Lebesgue set of f.

25. If E is a Borel set in $\mathbf{R}^n$, the **density** $D_E(x)$ of E at x is defined as

$$D_E(x) = \lim_{r \to 0} \frac{m(E \cap B(r,x))}{m(B(r,x))},$$

provided the limit exists.

(a) Show that $D_E(x) = 1$ for a.e. $x \in E$ and $D_E(x) = 0$ for a.e. $x \in E^c$.

(b) Find examples of E, x such that $D_E(x) = \alpha$, where $0 < \alpha < 1$, or such that $D_E(x)$ does not exist.

26. If $f \in L^1(m)$, or if f is $\overline{\mathbf{R}}$-valued and $f \in L^1_{loc}$, then the complex or signed measure $d\nu = f\,dm$ is regular. [Use Theorem (2.40) and Corollary (3.6).]

5. FUNCTIONS OF BOUNDED VARIATION

The theorems of the preceding section apply in particular on $\mathbf{R}^1$, where, because of the correspondence between regular Borel measures and increasing functions which we established in Section 1.5, they yield results about differentiation and integration of functions. As in Section 1.5, we adopt the notation that if F is an increasing, right continuous function on $\mathbf{R}$, μ_F is the Borel measure determined by the relation $\mu_F((a, b]) = F(b) - F(a)$. Also, throughout this section the term "almost everywhere" will always refer to Lebesgue measure.

(3.23) Theorem. Let F: $\mathbf{R} \to \mathbf{R}$ be increasing, and let $G(x) = F(x^+)$.

(a) The set of points at which F is discontinuous is countable.

(b) F and G are differentiable a.e., and $F' = G'$ a.e.

Proof. (a) Since F is increasing, the intervals $(F(x^-), F(x^+))$ $(x \in \mathbf{R})$ are disjoint, and for $|x| < N$ they lie in the interval $(F(-N), F(N))$. Hence

$$\sum_{|x|<N} [F(x^+) - F(x^-)] \leq F(N) - F(-N) < \infty,$$

which implies that $\{x \in (-N, N)\colon F(x^+) \neq F(x^-)\}$ is countable. As this is true for all N, (a) is proved.

(b) G is increasing and right continuous, and $G = F$ except perhaps where F is discontinuous. Applying Theorem (3.23) to μ_G and the families $\{(x - r, x]\}$ and $\{(x, x + r]\}$, which shrink nicely to x, and setting $h = \pm r$, we see that

$$\lim_{h \to 0} h^{-1}[G(x + h) - G(x)] = G'(x)$$

exists for a.e. x. Thus, if $H = G - F$, it suffices to show that H' exists and equals zero a.e.

Let $\{x_j\}$ be an enumeration of the points at which $H \neq 0$. Then $H(x_j) > 0$, and as in the proof of (a) we see that $\sum_{J_N} H(x_j) < \infty$ where $J_N = \{j\colon |x_j| < N\}$. Let δ_j be the point mass at x_j and let $\mu = \sum_j H(x_j)\delta_j$. By the preceding remark, μ is a regular Borel measure, and obviously $\mu \perp m$, so by Theorem (3.23), $r^{-1}\mu((x - r, x + r)) \to 0$ as $r \to 0$ for a.e. x, say $x \in E$. But if $h \in \mathbf{R}$ and we set $r = 2|h|$, we have

$$\left|\frac{H(x + h) - H(x)}{h}\right| \leq \frac{2[H(x + h) + H(x)]}{r}$$

$$\leq \frac{2\mu((x - r, x + r))}{r} \to 0 \quad \text{as } h \to 0$$

for $x \in E$, and hence $H' = 0$ on E. □

To proceed further we need some new terminology. If $F\colon \mathbf{R} \to \mathbf{C}$ and $x \in \mathbf{R}$, we define

$$T_F(x) = \sup\left\{\sum_1^n |F(x_j) - F(x_{j-1})|\colon \right.$$

$$\left. -\infty < x_0 < x_1 < \cdots < x_n = x, n \in \mathbf{N}\right\}.$$

T_F is called the **total variation function** of F. We observe that the sums in the definition of T_F are made larger if the number of subdivision points x_j is increased, so if $a < b$, the definition of $T_F(b)$ is unaffected if we assume that a is always one of the subdivision points. It follows that

(3.24)

$$T_F(b) = T_F(a) + \sup\left\{\sum_1^n |F(x_j) - F(x_{j-1})|\colon a = x_0 < \cdots < x_n = b\right\}.$$

Thus T_F is an increasing function with values in $[0, \infty]$. If $T_F(\infty) = \lim_{x \to \infty} T_F(x)$ is finite, we say that F is of **bounded variation** (on $\mathbf{R}$), and we denote the space of all such F by BV.

More generally, the supremum on the right of (3.24) is called the **total variation** of F on $[a, b]$. It depends only on the values of F on $[a, b]$, so we may define $BV([a, b])$ to be the set of all functions on $[a, b]$ whose total variation on $[a, b]$ is finite. If $F \in BV$, the restriction of F to $[a, b]$ is in $BV([a, b])$ for all a, b; indeed, its total variation on $[a, b]$ is nothing but $T_F(b) - T_F(a)$. Conversely, if $F \in BV([a, b])$ and we set $F(x) = F(a)$ for $x < a$ and $F(x) = F(b)$ for $x > b$, then $F \in BV$. By this device the results which we shall prove for BV can also be applied to $BV([a, b])$.

(3.25) Examples

(a) If $F\colon \mathbf{R} \to \mathbf{R}$ is bounded and increasing, then $F \in BV$ [in fact, $T_F(x) = F(x) - F(-\infty)$].

(b) If $F, G \in BV$ and $a, b \in \mathbf{C}$, then $aF + bG \in BV$.

(c) If F is differentiable on $\boldsymbol{R}$ and F' is bounded, then $F \in BV([a, b])$ for $-\infty < a < b < \infty$ (by the mean value theorem).

(d) If $F(x) = \sin x$, then $F \in BV([a, b])$ for $-\infty < a < b < \infty$ but $F \notin BV$.

(e) If $F(x) = x \sin(x^{-1})$ for $x \neq 0$ and $F(0) = 0$, then $F \notin BV([a, b])$ for $a \leq 0 < b$ or $a < 0 \leq b$.

The verification of these examples is left to the reader (Exercise 27).

(3.26) Lemma. If $F \in BV$ is real-valued, then $T_F + F$ and $T_F - F$ are increasing.

Proof. If $x < y$ and $\varepsilon > 0$, choose $x_0 < \cdots < x_n = x$ such that

$$\sum_1^n |F(x_j) - F(x_{j-1})| \geq T_F(x) - \varepsilon.$$

Then

$$T_F(y) + F(y) \geq \sum_1^n |F(x_j) - F(x_{j-1})| + |F(y) - F(x)|$$

$$+ (F(y) - F(x)) + F(x)$$

$$\geq T_F(x) + F(x) - \varepsilon.$$

Since ε is arbitrary, $T_F + F$ is increasing; likewise for $T_F - F$. □

(3.27) Theorem. (a) $F \in BV$ iff $\operatorname{Re} F \in BV$ and $\operatorname{Im} F \in BV$.

(b) If F: $\mathbf{R} \to \mathbf{R}$, then $F \in BV$ iff F is the difference of two bounded increasing functions; for $F \in BV$ these functions may be taken to be $\frac{1}{2}(T_F + F)$ and $\frac{1}{2}(T_F - F)$.

(c) If $F \in BV$, then $F(x^+) = \lim_{y \to x^+} F(y)$ and $F(x^-) = \lim_{y \to x^-} F(y)$ exist for all $x \in \mathbf{R}$, as do $F(\pm\infty) = \lim_{y \to \pm\infty} F(y)$.

(d) If $F \in BV$, the set of points of discontinuity of F is countable.

(e) If $F \in BV$ and $G(x) = F(x^+)$, then F' and G' exist and are equal a.e.

Proof. (a) is obvious. For (b), the "if" implication is easy [cf. (3.25a, b)]. To prove "only if," observe that by Lemma (3.26), the equation $F = \frac{1}{2}(T_F + F) - \frac{1}{2}(T_F - F)$ expresses F as the difference of two increasing functions. Also, the inequalities

$$T_F(y) \pm F(y) \geq T_F(x) \pm F(x) \quad \text{for } y > x$$

imply that

$$|F(y) - F(x)| \leq T_F(y) - T_F(x) \leq T_F(\infty) - T_F(-\infty) < \infty,$$

so F, and hence $T_F \pm F$, is bounded. Finally (c), (d), and (e) follow from (a), (b), and Theorem (3.23). □

The representation $F = \frac{1}{2}(T_F + F) - \frac{1}{2}(T_F - F)$ of a real-valued $F \in BV$ is called the **Jordan decomposition** of F, and $\frac{1}{2}(T_F + F)$ and $\frac{1}{2}(T_F - F)$ are called the **positive and negative variations** of F. Since for $x \in \mathbf{R}$, we have $x^+ = \max(x, 0) = \frac{1}{2}(|x| + x)$ and $x^- = \max(-x, 0) = \frac{1}{2}(|x| - x)$, it follows that

$$\tfrac{1}{2}(T_F + F)(x) = \sup\left\{ \sum_1^n \left[F(x_j) - F(x_{j-1}) \right]^+ : \right.$$
$$\left. -\infty < x_0 < \cdots < x_n = x \right\} + F(-\infty),$$

$$\tfrac{1}{2}(T_F - F)(x) = \sup\left\{ \sum_1^n \left[F(x_j) - F(x_{j-1}) \right]^- : \right.$$
$$\left. -\infty < x_0 < \cdots < x_n = x \right\} - F(-\infty).$$

Theorem (3.27a, b) suggests a close connection between BV and the space of complex Borel measures on $\mathbf{R}$. To make this precise, let NBV (N for "normalized") denote the space of all $F \in BV$ such that $F(-\infty) = 0$ and F is

right continuous. We observe that if $F \in BV$ then the function G defined by $G(x) = F(x^+) - F(-\infty)$ is in NBV and $G' = F'$ a.e. [That $G \in BV$ follows easily from Theorem (3.27): if $F = F_1 - F_2$ where F_1, F_2 are increasing, then $G(x) = F_1(x^+) - [F_2(x^+) + F(-\infty)]$, which is again the difference of two increasing functions.]

(3.28) Lemma. If $F \in BV$, then $T_F(-\infty) = 0$. If also F is right continuous, then so is T_F.

Proof. If $\varepsilon > 0$ and $x \in \mathbf{R}$, choose $x_0 < \cdots < x_n = x$ so that

$$\sum_1^n |F(x_j) - F(x_{j-1})| \geq T_F(x) - \varepsilon.$$

From (3.24) we see that $T_F(x) - T_F(x_0) \geq T_F(x) - \varepsilon$, and hence $T_F(y) \leq \varepsilon$ for $y \leq x_0$. Thus $T_F(-\infty) = 0$. If F is right continuous and $x \in \mathbf{R}$, let $\alpha = T_F(x^+) - T_F(x)$. Given $\varepsilon > 0$ we can choose $\delta > 0$ so that if $0 < y < \delta$ then $|F(x + y) - F(x)| < \varepsilon$ and $T_F(x + y) - T_F(x^+) < \varepsilon$. For any such y, by (3.24) there exist $x = x_0 < \cdots < x_n = x + y$ such that $\sum_1^n |F(x_j) - F(x_{j-1})| \geq 3\alpha/4$, and hence $\sum_2^n |F(x_j) - F(x_{j-1})| \geq (3\alpha/4) - \varepsilon$. But then there exist $x = t_0 < \cdots < t_m = x_1$ such that $\sum_1^m |F(t_j) - F(t_{j-1})| \geq 3\alpha/4$. Hence

$$\alpha + \varepsilon > T_F(x + y) - T_F(x)$$
$$\geq \sum_1^m |F(t_j) - F(t_{j-1})| + \sum_2^n |F(x_j) - F(x_{j-1})| \geq (3\alpha/2) - \varepsilon.$$

Thus $\alpha < 4\varepsilon$, and since ε is arbitrary, $\alpha = 0$. □

(3.29) Theorem. If μ is a complex Borel measure on $\mathbf{R}$ and $F(x) = \mu((-\infty, x])$, then $F \in NBV$. Conversely, if $F \in NBV$, there is a unique complex Borel measure μ_F such that $F(x) = \mu_F((-\infty, x])$; moreover, $|\mu_F| = \mu_{T_F}$.

Proof. If μ is a complex measure, we have $\mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-)$ where $\mu_j^\pm$ are finite measures. If $F_j^\pm(x) = \mu_j^\pm((-\infty, x])$, then $F_j^\pm$ is increasing and right continuous, $F_j^\pm(-\infty) = 0$, and $F_j^\pm(\infty) = \mu_j^\pm(\mathbf{R}) < \infty$. By Theorem (3.27), $F = F_1^+ - F_1^- + i(F_2^+ - F_2^-) \in NBV$. Conversely, by Theorem (3.27) and Lemma (3.28) any $F \in NBV$ can be written in this form, and each $F_j^\pm$ gives rise to a measure $\mu_j^\pm$, so $F(x) = \mu_F((-\infty, x])$ where $\mu_F = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-)$. The proof that $|\mu_F| = \mu_{T_F}$ is outlined in Exercise 28. □

If $F \in NBV$, it is customary to denote the integral of a function g with respect to the measure μ_F by $\int g dF$ or $\int g(x) dF(x)$; such integrals are called

Lebesgue–Stieltjes integrals. At this point we present an integration-by-parts formula for Lebesgue–Stieltjes integrals; other variants if this result can be found in Exercises 34 and 35.

(3.30) Theorem. If F and G are in NBV and at least one of them is continuous, then for $-\infty < a < b < \infty$,

$$\int_{(a,b]} F\,dG + \int_{(a,b]} G\,dF = F(b)G(b) - F(a)G(a).$$

Proof. F and G are linear combinations of increasing functions in NBV, so a simple calculation shows that it suffices to assume F and G increasing. Suppose for the sake of definiteness that G is continuous, and let $\Delta = \{(x, y): a < x \le y \le b\}$. We use Fubini's theorem to compute $\mu_F \times \mu_G(\Delta)$ in two ways:

$$\mu_F \times \mu_G(\Delta) = \int_{(a,b]}\int_{(a,y]} dF(x)\,dG(y) = \int_{(a,b]} [F(y) - F(a)]\,dG(y)$$

$$= \int_{(a,b]} F\,dG - F(a)[G(b) - G(a)],$$

and since $G(x) = G(x^-)$,

$$\mu_F \times \mu_G(\Delta) = \int_{(a,b]}\int_{[x,b]} dG(y)\,dF(x) = \int_{(a,b]} [G(b) - G(x)]\,dF(x)$$

$$= G(b)[F(b) - F(a)] - \int_{(a,b]} G\,dF.$$

Subtracting these two equations, we obtain the desired result. □

The next obvious question is: Which functions in NBV correspond to measures μ such that $\mu \perp m$ or $\mu \ll m$? One answer is the following:

(3.31) Proposition. If $F \in NBV$, then $F' \in L^1(m)$. Moreover, $\mu_F \perp m$ iff $F' = 0$ a.e., and $\mu_F \ll m$ iff $F(x) = \int_{-\infty}^{x} F'(t)\,dt$.

Proof. We have merely to observe that $F'(x) = \lim_{r\to 0} \mu_F(E_r)/m(E_r)$, where $E_r = (x, x + r]$ or $(x - r, x]$, and apply Theorem (3.22). □

The condition $\mu_F \ll m$ can also be expressed directly in terms of F, as follows. A function F: $\mathbf{R} \to \mathbf{C}$ is called **absolutely continuous** if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for any finite set of disjoint intervals

$(a_1, b_1), \ldots, (a_N, b_N)$,

$$\text{(3.32)} \qquad \sum_1^N (b_j - a_j) < \delta \text{ implies } \sum_1^N |F(b_j) - F(a_j)| < \varepsilon.$$

More generally, F is said to be **absolutely continuous** on $[a, b]$ if this condition holds when the intervals (a_j, b_j) are restricted to lie in $[a, b]$. Clearly, if F is absolutely continuous, then F is uniformly continuous [take $N = 1$ in (3.32)], and if F is everywhere differentiable and F' is bounded, then F is absolutely continuous (by the mean value theorem).

(3.33) Proposition. If $F \in NBV$, then F is absolutely continuous iff $\mu_F \ll m$.

Proof. If $\mu_F \ll m$, the absolute continuity of F follows by applying Theorem (3.5) to the sets $E = \bigcup_1^N (a_j, b_j]$. To prove the converse, suppose that E is a Borel set such that $m(E) = 0$. If ε and δ are as in the definition of absolute continuity of F, by Theorem (1.18) we can find open sets $U_1 \supset U_2 \supset \cdots \supset E$ such that $m(U_1) < \delta$ [and thus $m(U_j) < \delta$ for all j] and $\mu_F(U_j) \to \mu_F(E)$. Each U_j is a disjoint union of open intervals (a_j^k, b_j^k), and

$$\sum_{k=1}^N |\mu_F((a_j^k, b_j^k))| \le \sum_{k=1}^N |F(b_j^k) - F(a_j^k)| < \varepsilon$$

for all N. Hence, letting $N \to \infty$, we obtain $|\mu_F(U_j)| < \varepsilon$ and hence $|\mu_F(E)| < \varepsilon$. Since ε is arbitrary, $\mu_F(E) = 0$, which shows that $\mu_F \ll m$. □

(3.34) Corollary. If $f \in L^1(m)$, then the function $F(x) = \int_{-\infty}^{x} f(t)\,dt$ is in NBV and is absolutely continuous, and $f = F'$ a.e. Conversely, if $F \in NBV$ is absolutely continuous, then $F' \in L^1(m)$ and $F(x) = \int_{-\infty}^{x} F'(t)\,dt$.

Proof. This follows immediately from Propositions (3.31) and (3.33). □

If we consider functions on bounded intervals, this result can be refined a bit.

(3.35) Lemma. If F is absolutely continuous on $[a, b]$, then $F \in BV([a, b])$.

Proof. Let δ be as in the definition of absolute continuity, corresponding to $\varepsilon = 1$. If $a = x_0 < \cdots < x_n = b$, by inserting more subdivision points if necessary we can collect the intervals (x_{j-1}, x_j) into at most N groups such that the sum of the lengths in each group is less than δ, where N is the greatest integer less than $\delta^{-1}(b - a) + 1$. It follows that the total variation of F on $[a, b]$ is at most N. □

(3.36) The Fundamental Theorem of Calculus for Lebesgue Integrals. If $-\infty < a < b < \infty$ and $F: [a, b] \to \mathbf{C}$, the following are equivalent:

(a) F is absolutely continuous on $[a, b]$.

(b) $F(x) - F(a) = \int_a^x f(t)\,dt$ for some $f \in L^1([a, b], m)$.

(c) F is differentiable a.e. on $[a, b]$, $F' \in L^1([a, b], m)$, and $F(x) - F(a) = \int_a^x F'(t)\,dt$.

Proof. To prove that (a) implies (c), we may assume without loss of generality that $F(a) = 0$. If we set $F(x) = 0$ for $x < a$ and $F(x) = F(b)$ for $x > b$, then $F \in NBV$ by Lemma (3.35), so (c) follows from Corollary (3.34). That (c) implies (b) is trivial. Finally, (b) implies (a) by setting $f(t) = 0$ for $t \notin [a, b]$ and applying Corollary (3.34). □

We conclude with some further remarks on decomposition of measures. A complex Borel measure μ on $\mathbf{R}^n$ is called **discrete** if there is a countable set $\{x_j\} \subset \mathbf{R}^n$ and complex numbers $\{c_j\}$ such that $\sum|c_j| < \infty$ and $\mu = \sum c_j \delta_{x_j}$ where δ_x is the point mass at x. On the other hand, μ is called **continuous** if $\mu(\{x\}) = 0$ for all $x \in \mathbf{R}^n$. Any complex Borel measure μ can be written uniquely as $\mu = \mu_d + \mu_c$ where μ_d is discrete and μ_c is continuous. Indeed, let $E = \{x: \mu(\{x\}) \neq 0\}$. For any countable subset F of E the series $\sum_{x \in F} \mu(\{x\})$ converges absolutely [to $\mu(F)$], so $\{x \in E: |\mu(\{x\})| > k^{-1}\}$ is finite for all k, and it follows that E itself is countable. Hence $\mu_d(A) = \mu(A \cap E)$ is discrete and $\mu_c(A) = \mu(A \setminus E)$ is continuous.

Obviously, if $\mu \ll m$ then μ is continuous. Thus by Theorem 3.22, any (regular) complex Borel measure μ on $\mathbf{R}^n$ can be written uniquely as $\mu = \mu_d + \mu_s + \mu_a$ where μ_d is discrete, μ_s is continuous and $\mu_s \perp m$, and $\mu_a \ll m$.

Applying this result in the one-dimensional case together with the results of this section, we see that any $F \in NBV$ can be written uniquely as $F = F_a + F_s + F_d$ where $F_a, F_s, F_d \in NBV$, F_a is absolutely continuous, F_s is continuous and $F_s' = 0$ a.e., and F_d is a "jump function": that is, there exist a sequence $\{x_j\} \subset \mathbf{R}$ and a sequence $\{c_j\} \subset \mathbf{C}$ such that $\sum|c_j| < \infty$ and $F(x) = \sum_{x_j \le x} c_j$. The Cantor function constructed in Section 1.5 is an example of a nontrivial function of type F_s. More surprisingly, there exist strictly increasing functions of type F_s: see Exercise 40.

EXERCISES

27. Verify the assertions in (3.25).

28. If $F \in NBV$, let $G(x) = |\mu_F|((-\infty, x])$.

(a) From the definition of T_F, $T_F \le G$.

(b) $|\mu_F(E)| \le \mu_{T_F}(E)$ when E is an interval, hence when E is a Borel set.

(c) $|\mu_F| \le \mu_{T_F}$ and hence $G \le T_F$. (Use Exercise 21.)

(d) $|\mu_F| = \mu_{T_F}$.

29. If $F \in NBV$ is real-valued then $\mu_F^+ = \mu_P$ and $\mu_F^- = \mu_N$ where P and N are the positive and negative variations of F. (Use Exercise 28.)

30. Construct an increasing function on $\mathbf{R}$ whose set of discontinuities is $\mathbf{Q}$.

31. Let $F(x) = x^2\sin(x^{-1})$ and $G(x) = x^2\sin(x^{-2})$ for $x \neq 0$, and $F(0) = G(0) = 0$.

(a) F and G are differentiable everywhere (including $x = 0$).

(b) $F \in BV([-1, 1])$ and $G \notin BV([-1, 1])$.

32. If $F_j, F \in NBV$ and $F_j \to F$ pointwise, then $T_F \le \liminf T_{F_j}$.

33. If F is increasing on $\mathbf{R}$, then $F(b) - F(a) \ge \int_a^b F'(t)\,dt$.

34. Suppose $F, G \in NBV$ and $-\infty < a < b < \infty$.

(a) By adapting the proof of Theorem (3.30), show that

$$\int_{[a,b]} \frac{F(x) + F(x^-)}{2}\,dG(x) + \int_{[a,b]} \frac{G(x) + G(x^-)}{2}\,dF(x)$$

$$= F(b)G(b) - F(a^-)G(a^-).$$

(b) If there are no points in $[a, b]$ where F and G are both discontinuous, then

$$\int_{[a,b]} F\,dG + \int_{[a,b]} G\,dF = F(b)G(b) - F(a^-)G(a^-).$$

35. If F and G are absolutely continuous on $[a, b]$, then so is FG, and

$$\int_a^b (FG' + F'G)(x)\,dx = F(b)G(b) - F(a)G(a).$$

36. Let G be a continuous increasing function on $[a, b]$ and let $G(a) = c$, $G(b) = d$.

(a) If $E \subset [c, d]$ is a Borel set, then $m(E) = \mu_G(G^{-1}(E))$. (First consider the case where E is an interval.)

(b) If f is Borel measurable and integrable on $[c, d]$, then $\int_c^d f(y)dy = \int_a^b f(G(x))dG(x)$. [In particular, if G is absolutely continuous, then $\int_c^d f(y)dy = \int_a^b f(G(x))G'(x)dx$.]

(c) Show by example that (b) may fail if G is merely right continuous rather than continuous.

37. Suppose $F: \mathbf{R} \to \mathbf{C}$. There is a constant M such that $|F(x) - F(y)| \le M|x - y|$ for all $x, y \in \mathbf{R}$ (that is, F is **Lipschitz continuous**) iff F is absolutely continuous and $|F'| \le M$ a.e.

38. If f: $[a, b] \to \mathbf{R}$, consider the graph of f as a subset of $\mathbf{C}$, namely $\{t + if(t)\colon\ t \in [a, b]\}$. The length L of this graph is by definition the supremum of the lengths of all inscribed polygons.

(a) If $F(t) = t + if(t)$, L is the total variation of F on $[a, b]$.

(b) If f is absolutely continuous, $L = \int_a^b [1 + f'(t)^2]^{1/2} dt$.

39. (Fubini) If $\{F_j\}$ is a sequence of increasing functions on $[a, b]$ such that $F(x) = \sum_1^\infty F_j(x) < \infty$ for all $x \in [a, b]$, then $F'(x) = \sum_1^\infty F_j'(x)$ for a.e. $x \in [a, b]$. (It suffices to assume $F_j \in NBV$. Consider the measures μ_{F_j}.)

40. Let F denote the Cantor function on $[0, 1]$, and set $F(x) = 0$ for $x < 0$, $F(x) = 1$ for $x > 1$. Let $\{[a_n, b_n]\}$ be an enumeration of the closed subintervals of $[0, 1]$ with rational endpoints, and let

$$F_n(x) = F\left(\frac{x - a_n}{b_n - a_n}\right).$$

Then $G = \sum_1^\infty 2^{-n} F_n$ is continuous and strictly increasing on $[0, 1]$, and $G' = 0$ a.e. (Use Exercise 39.)

41. (a) There exists a Borel set $A \subset [0, 1]$ such that $0 < m(A \cap I) < m(I)$ for every subinterval I of $[0, 1]$. (Hint: every subinterval of $[0, 1]$ contains closed, nowhere dense subsets of positive measure.)

(b) Let $F(x) = m([0, x] \cap A)$. Then F is absolutely continuous and strictly increasing on $[0, 1]$, but $F' = 0$ on a set of positive measure.

(c) Let $G(x) = m([0, x] \cap A) - m([0, x] \setminus A)$. Then G is absolutely continuous on $[0, 1]$, but G is not monotone on any subinterval of $[0, 1]$.

42. A function F: $(a, b) \to \mathbf{R}$ $(-\infty \le a < b \le \infty)$ is called **convex** if $F(\lambda s + (1 - \lambda)t) \le \lambda F(s) + (1 - \lambda)F(t)$ for all $s, t \in (a, b)$ and $\lambda \in (0, 1)$.

(a) F is convex iff for all s, t, s', t' in (a, b) such that $s \le s' < t'$ and $s < t \le t'$,

$$\frac{F(t) - F(s)}{t - s} \le \frac{F(t') - F(s')}{t' - s'}.$$

(b) F is convex on (a, b) iff F is absolutely continuous on every compact subinterval of (a, b) and F' is increasing (on the set where it is defined).

(c) If F is convex on (a, b) and $t_0 \in (a, b)$, there exists $\beta \in \mathbf{R}$ such that $F(t) - F(t_0) \ge \beta(t - t_0)$ for all $t \in (a, b)$.

(d) (Jensen's inequality) If $(X, \mathcal{M}, \mu)$ is a measure space with $\mu(X) = 1$, g: $X \to (a, b)$ is in $L^1(\mu)$, and F is convex on (a, b), then $F(\int g\, d\mu) \le \int F \circ g\, d\mu$. [Let $t_0 = \int g\, d\mu$, $t = g(x)$. Apply part (c) and integrate.]

6. NOTES AND REFERENCES

Section 1. Our proof of the Hahn decomposition theorem follows Doss [24].

Section 2. The Lebesgue–Radon–Nikodym theorem was proved by Lebesgue [63] in the case where μ is Lebesgue measure on $\mathbf{R}^n$. Under the hypothesis $\nu \ll \mu$, it was generalized by Radon [77] to arbitrary regular Borel measures on $\mathbf{R}^n$ and by Nikodym [75] to measures on abstract spaces. The Lebesgue decomposition in the abstract setting appears in Saks [93]. The proof of the Lebesgue–Radon–Nikodym theorem in the text is similar to, but more efficient than, the one in Halmos [41]; I learned it from L. Loomis.

Section 3. The characterization

$$|\nu|(E) = \sup\left\{\sum_1^n |\nu(E_j)|\colon\ E_1, \ldots, E_n \text{ disjoint},\ E = \bigcup_1^n E_j\right\}$$

of the total variation of a complex measure ν (cf. Exercise 21) is usually taken as the definition of $|\nu|$. Our definition seems more generally useful, and it is certainly easier to compute with.

Section 4. Theorems (3.21) and (3.22) are due to Lebesgue [63], but the line of argument we have presented is essentially that of Wiener [110]. Our proof of Theorem (3.18) is illustrative of a general technique that has been much exploited in recent years, namely, controlling the limiting behavior of a family of operators by means of estimates on an appropriate maximal function.

Lemma (3.15), a simplified version of Wiener's covering lemma, is taken from Rudin [90]. There is also an older and more delicate covering theorem, due to Vitali, which is used for similar purposes:

If $E \subset \mathbf{R}^n$ and $\mathcal{Q}$ is a family of cubes such that each $x \in E$ is contained in members of $\mathcal{Q}$ of arbitrarily small diameter, there is a (finite or infinite) disjoint sequence $\{Q_j\} \subset \mathcal{Q}$ such that $m(E \setminus \bigcup_j Q_j) = 0$.

Proofs can be found in many books, for example, Saks [93] and Cohn [19].

The maximal function Hf, in dimension one, was first studied in Hardy and Littlewood [44]. This paper is notable for (among other things) its explanation of nonincreasing rearrangements in terms of cricket averages.

Section 5. The main results of this section are due to Lebesgue and Vitali; see Hawkins [49] for detailed references. Exercise 36 gives one form of the change-of-variable formula for Lebesgue integrals; others can be found in

Serrin and Varberg [97]. The example in Exercise 40 is due to Brown [15]. For a terse solution of Exercise 41a, see Rudin [92].

The example of the Cantor function shows that a continuous, almost everywhere differentiable function need not be the integral of its derivative. However, one can ask: if F: $[a, b] \to \mathbf{R}$ is differentiable at *every* $x \in [a, b]$, can F be recovered from F' by integration? The answer is *yes* provided that $F' \in L^1$; in fact, one can then allow F' to be undefined at countably many points. A proof of this highly nontrivial theorem can be found in Cohn [19], Theorem 6.3.10; see also Rudin [90], Theorem 8.21, for the somewhat simpler case when F' is assumed to exist everywhere. However, there exist everywhere differentiable functions F such that $F' \notin L^1$. The simplest example is $F(x) = x^2 \sin(x^{-2})$ (cf. Exercise 31), but it is not hard to construct examples in which the singularities of F' are much more complicated. To integrate such an F' one must extend the Lebesgue integral to include "conditionally convergent" integrals of functions f such that $\int f^+ = \int f^- = \infty$. Such extended integrals, whose domains include all derivatives of everywhere differentiable functions, have been constructed by Denjoy, Perron, and others; accounts of them may be found in Saks [93] and Lebesgue [64].

4

POINT SET TOPOLOGY

The concepts of limit, convergence, and continuity are central to all of analysis, and it is useful to have a general framework for studying them which includes their classical manifestations as special cases. One such framework, which has the advantage of not requiring many ideas beyond those occurring in analysis on Euclidean space, is that of metric spaces. However, metric spaces are not sufficiently general to describe even some very classical modes of convergence, for example, pointwise convergence of functions on **R**. A more flexible theory can be built by taking the open sets, rather than a metric, as the primitive data, and it is this theory that we shall explore in the present chapter.

1. TOPOLOGICAL SPACES

Let X be a nonempty set. A **topology** on X is a family $\mathcal{T}$ of subsets of X which contains $\varnothing$ and X and is closed under arbitrary unions and finite intersections (i.e., if $\{U_\alpha\}_{\alpha \in A} \subset \mathcal{T}$ then $\bigcup_{\alpha \in A} U_\alpha \in \mathcal{T}$, and if $U_1, \ldots, U_n \in \mathcal{T}$ then $\bigcap_1^n U_j \in \mathcal{T}$). The pair $(X, \mathcal{T})$ is called a **topological space**. If $\mathcal{T}$ is understood, we shall simply refer to the topological space X. Let us examine a few examples:

(i) If X is any nonempty set, $\mathcal{P}(X)$ and $\{\varnothing, X\}$ are topologies on X. They are called the **discrete topology** and the **trivial topology**, respectively.

(ii) If X is an infinite set, $\{U \subset S: U = \varnothing \text{ or } U^c \text{ is finite}\}$ is a topology on X, called the **confinite topology**.

(iii) If X is a metric space, the collection of all open sets with respect to the metric is a topology on X.

(iv) If $(X, \mathcal{T})$ is a topological space and $Y \subset X$, then $\mathcal{T}_Y = \{U \cap Y: U \in \mathcal{T}\}$ is a topology on Y, called the **relative topology** induced by $\mathcal{T}$.

We now present the basic terminology concerning topological spaces. Most of these concepts are already familiar in the context of metric spaces [example (iii)]. Until further notice, $(X, \mathscr{T})$ will be a fixed topological space.

The elements of $\mathscr{T}$ are called **open sets,** and their complements are called **closed sets**. If $Y \subset X$, the open (closed) subsets of Y in the relative topology are called **relatively** open (closed). We observe that, by deMorgan's laws, the family of closed sets is closed under arbitrary intersections and finite unions.

If $A \subset X$, the union of all open sets contained in A is called the **interior** of A, and the intersection of all closed sets containing A is called the **closure** of A. We denote the interior and closure of A by A^o and $\bar{A}$. Clearly A^o is the largest open set contained in A and $\bar{A}$ is the smallest closed set containing A, and we have $(A^o)^c = \overline{A^c}$ and $(\bar{A})^c = (A^c)^o$. The difference $\bar{A} \setminus A^o = \bar{A} \cap \overline{A^c}$ is called the **boundary** of A and is denoted by ∂A. If $\bar{A} = X$, A is called **dense** in X. On the other hand, if $(\bar{A})^o = \varnothing$, A is called **nowhere dense.**

If $x \in X$ (or $E \subset X$), a **neighborhood** of x (or E) is a set $A \subset X$ such that $x \in A^o$ (or $E \subset A^o$). Thus, a set A is open iff it is a neighborhood of itself. (Some authors require neighborhoods to be open sets; we do not.) If $x \in X$ and $A \subset X$, x is called an **accumulation point** of A if $A \cap (U \setminus \{x\}) \neq \varnothing$ for every neighborhood U of x. (Other terms sometimes used for the same concept are "cluster point" and "limit point." We shall use "cluster point" to mean something a bit different below.)

(4.1) Proposition. If $A \subset X$, let $C(A)$ be the set of accumulation points of A. Then $\bar{A} = A \cup C(A)$, and A is closed iff $C(A) \subset A$.

Proof. If $x \notin \bar{A}$, then A^c is a neighborhood of x which does not intersect A, so $x \notin C(A)$; thus $A \cup C(A) \subset \bar{A}$. If $x \notin C(A) \cup A$, there is an open U containing x such that $U \cap A = \varnothing$, and hence $\bar{A} \subset U^c$ and $x \notin \bar{A}$. Thus $\bar{A} \subset C(A) \cup A$. Finally, A is closed iff $A = \bar{A}$, and this happens iff $C(A) \subset A$. □

If $\mathscr{T}_1$ and $\mathscr{T}_2$ are topologies on X such that $\mathscr{T}_1 \subset \mathscr{T}_2$, we say that $\mathscr{T}_1$ is **weaker** (or **coarser**) than $\mathscr{T}_2$, or that $\mathscr{T}_2$ is **stronger** (or **finer**) than $\mathscr{T}_1$. Clearly the trivial topology is the weakest topology on X, while the discrete topology is the strongest. If $\mathscr{E} \subset \mathscr{P}(X)$, there is a unique weakest topology $\mathscr{T}(\mathscr{E})$ containing $\mathscr{E}$, namely the intersection of all topologies containing $\mathscr{E}$. It is called the topology **generated** by $\mathscr{E}$, and $\mathscr{E}$ is sometimes called a **subbase** for $\mathscr{T}(\mathscr{E})$.

If $\mathscr{T}$ is a topology on X, a **neighborhood base** for $\mathscr{T}$ at $x \in X$ is a family $\mathscr{N} \subset \mathscr{T}$ such that if $U \in \mathscr{T}$ and $x \in U$ there exists $V \in \mathscr{N}$ such that $x \in V$ and $V \subset U$. A **base** for $\mathscr{T}$ is a family $\mathscr{B} \subset \mathscr{T}$ which contains a neighborhood base for $\mathscr{T}$ at every $x \in X$. For example, if X is a metric space, the collection of open balls centered at x is a neighborhood base for the metric topology at x, and the collection of all open balls in X is a base.

(4.2) Proposition. If $\mathscr{T}$ is a topology on X and $\mathscr{E} \subset \mathscr{T}$, then $\mathscr{E}$ is a base for $\mathscr{T}$ iff every nonempty $U \in \mathscr{T}$ is a union of sets in $\mathscr{E}$.

Proof. If $\mathscr{E}$ is a base, $U \in \mathscr{T}$, and $x \in U$, there exists $V_x \in \mathscr{E}$ with $x \in V_x \subset U$, so $U = \bigcup_{x \in U} V_x$. The reverse implication is obvious. □

(4.3) Proposition. If $\mathscr{E} \subset \mathscr{P}(X)$, in order for $\mathscr{E}$ to be a base for a topology on X it is necessary and sufficient for the following two conditions to be satisfied: (a) each $x \in X$ is contained in some $V \in \mathscr{E}$, and (b) if $U, V \in \mathscr{E}$ and $x \in U \cap V$ there exists $W \in \mathscr{E}$ with $x \in W \subset (U \cap V)$.

Proof. The necessity is clear, since if U, V are open then so is $U \cap V$. To prove the sufficiency, let $\mathscr{T} = \{U \subset X$: for every $x \in U$ there exists $V \in \mathscr{E}$ with $x \in V \subset U\}$. Then $\varnothing \in \mathscr{T}$ and $X \in \mathscr{T}$ by condition (a), and $\mathscr{T}$ is obviously closed under unions. If $U_1, U_2 \in \mathscr{T}$ and $x \in U_1 \cap U_2$, there exist $V_1, V_2 \in \mathscr{E}$ with $x \in V_1 \subset U_1$, $x \in V_2 \subset U_2$, and by condition (b) there exists $W \in \mathscr{E}$ with $x \in W \subset (V_1 \cap V_2)$. Thus $U_1 \cap U_2 \in \mathscr{T}$, so by induction $\mathscr{T}$ is closed under finite intersections. $\mathscr{T}$ is therefore a topology, and $\mathscr{E}$ is clearly a base for $\mathscr{T}$. □

(4.4) Proposition. If $\mathscr{E} \subset \mathscr{P}(X)$, the topology $\mathscr{T}$ generated by $\mathscr{E}$ consists of $\varnothing$, X, and all unions of finite intersections of members of $\mathscr{E}$.

Proof. The family of finite intersections of sets in $\mathscr{E}$, together with X, satisfies the conditions of Proposition (4.3), so by Proposition (4.2) the family of all unions of such sets, together with $\varnothing$, is a topology. It is obviously contained in $\mathscr{T}$, hence equal to $\mathscr{T}$. □

Note how the simplicity of this proposition contrasts with the corresponding result for σ-algebras [Proposition (1.23)]. What makes life easy here is that only *finite* intersections are involved.

The concept of topological space is general enough to include a great profusion of interesting examples but—by the same token—too general to yield many interesting theorems. To build a reasonable theory one must usually restrict the class of spaces under consideration. The remainder of this section is devoted to a discussion of two types of restrictions that are commonly made: the so-called countability and separation axioms.

A topological space $(X, \mathscr{T})$ satisfies the **first axiom of countability,** or is **first countable**, if there is a countable neighborhood base for $\mathscr{T}$ at every $x \in X$. (It is useful to observe that if X is first countable, for any $x \in X$ there is a neighborhood base $\{U_j\}_1^\infty$ at x such that $U_j \supset U_{j+1}$ for all j. Indeed, if $\{V_i\}_1^\infty$ is any countable neighborhood base at x we can take $U_j = \bigcap_1^j V_i$.) $(X, \mathscr{T})$ satisfies the **second axiom of countability,** or is **second countable**, if $\mathscr{T}$ has a countable base. Every metric space is first countable (the balls of rational radius about x are a neighborhood base at x), and a metric space is second countable iff it is separable (Exercise 5). This fact can be partly generalized: a topological space, like a metric space, is called **separable** if it has a countable dense subset.

(4.5) Proposition. Every second countable space is separable.

Proof. If X is second countable, let $\mathscr{E}$ be a countable base for the topology, and for each $U \in \mathscr{E}$ pick a point $x_U \in U$. Then the complement of the closure of $\{x_U: U \in \mathscr{E}\}$ is an open set which does not include any $U \in \mathscr{E}$; hence it is empty and $\{x_U: U \in \mathscr{E}\}$ is dense. □

A sequence $\{x_j\}$ in a topological space X is said to **converge** to $x \in X$ (in symbols: $x_j \to x$) if for every neighborhood U of x there exists $J \in \mathbf{N}$ such that $x_j \in U$ for all $j > J$. First countable spaces have the pleasant property that such things as closure and continuity can be characterized in terms of sequential convergence—which is not the case in more general spaces, as we shall see. For example:

(4.6) Proposition. If X is first countable and $A \subset X$, then $x \in \bar{A}$ iff there is a sequence $\{x_j\}$ in A which converges to x.

Proof. Let $\{U_j\}$ be a countable neighborhood base at x with $U_j \supset U_{j+1}$ for all j. If $x \in \bar{A}$, then $U_j \cap A \neq \varnothing$ for all j. Pick $x_j \in U_j \cap A$: since $U_j \subset U_k$ for $j > k$ and every neighborhood of x contains some U_j, it is clear that $x_j \to x$. On the other hand, if $x \notin \bar{A}$ and $\{x_j\}$ is any sequence in A, then $(\bar{A})^c$ is a neighborhood of x containing no x_j, so $x_j \not\to x$. □

Finally, we discuss the separation axioms. These are properties of a topological space, labeled $T_0, \ldots, T_4$, which guarantee the existence of open sets which separate points or closed sets from each other. If X has property T_j, we say that X is a T_j space or that the topology on X is T_j.

T_0: If $x \neq y$, there is an open set containing x but not y *or* an open set containing y but not x.

T_1: If $x \neq y$, there is an open set containing y but not x.

T_2: If $x \neq y$, there are disjoint open sets U, V with $x \in U$ and $y \in V$.

T_3: X is a T_1 space, and for any closed set $A \subset X$ and any $x \in X \setminus A$ there are disjoint open sets U, V with $x \in U$ and $A \subset V$.

T_4: X is a T_1 space, and for any disjoint closed sets A, B in X there are disjoint open sets U, V with $A \subset U$ and $B \subset V$.

T_2, T_3, and T_4 also have other names: a T_2 space is a **Hausdorff** space, a T_3 space is a **regular** space, and a T_4 space is a **normal** space. (Some authors do not require regular and normal spaces to be T_1. There is an additional separation condition intermediate between T_3 and T_4 which we shall discuss in Section 2.) The following characterization of T_1 spaces is useful; it shows in particular that every normal space is regular and every regular space is Hausdorff.

(4.7) Proposition. X is a T_1 space iff $\{x\}$ is closed for every $x \in X$.

Proof. If X is T_1 and $x \in X$, for each $y \neq x$ there is an open U_y containing y but not x; thus $\{x\}^c = \bigcup_{y \neq x} U_y$ is open and $\{x\}$ is closed. Conversely, if $\{x\}$ is closed and $y \neq x$, then $\{x\}^c$ is an open set containing y but not x. □

The vast majority to topological spaces that arise in practice (and, in particular, in this book) are Hausdorff, or become Hausdorff after simple modifications. (This last phrase refers to spaces such as the space of integrable functions on a measure space, with the L^1 metric, which becomes Hausdorff when we identify two functions which are equal a.e.) However, one example—or, rather, class of examples—of non-Hausdorff topologies is of sufficient importance to warrant special mention: the Zariski topology on an algebraic variety. Without attempting to give the general definition of an algebraic variety, we shall describe the Zariski topology on a vector space.

Let k be a field, and let $k(X_1, \ldots, X_n)$ be the ring of polynomials in n variables over k. Each $P \in k(X_1, \ldots, X_n)$ determines a polynomial map p: $k^n \to k$ by substituting elements of k for the formal indeterminates $X_1, \ldots, X_n$. The correspondence $P \to p$ is one-to-one precisely when k is infinite. The collection of all sets $p^{-1}(\{0\})$ in k^n, as p ranges over all polynomial maps, is closed under finite unions, since $p^{-1}(\{0\}) \cap q^{-1}(\{0\}) = (pq)^{-1}(\{0\})$, and it contains k^n itself (take $p = 0$). It therefore follows from Propositions (4.2) and (4.3) that the collection of all sets of the form $\bigcap_{\alpha \in A} p_\alpha^{-1}(\{0\})$ (p_α being a polynomial map for each α) is the collection of closed sets for a topology on k^n, called the **Zariski topology**. The Zariski topology is T_1 by Proposition (4.7), for if $a = (a_1, \ldots, a_n) \in k^n$ then $\{a\} = \bigcap_1^n p_j^{-1}(\{0\})$ where $p_j(x_1, \ldots, x_n) = x_j - a_j$. If k is finite the Zariski topology is discrete, but if k is infinite the Zariski topology is not Hausdorff; in fact, any two nonempty open sets have nonempty intersection. This is just a restatement of the fact that $k(X_1, \ldots, X_n)$ is an integral domain, that is, if P and Q are nonzero polynomials then PQ is nonzero. (For $n = 1$, the Zariski topology on k is the cofinite topology.)

Other examples illustrating the separation and countability axioms will be found in the exercises.

EXERCISES

1. If $\text{card}(X) \geq 2$, there is a topology on X which is T_0 but not T_1.
2. If X is an infinite set, the cofinite topology on X is T_1 but not T_2, and is first countable iff X is countable.
3. Every metric space is normal. [If A, B are closed sets in the metric space (X, ρ), consider the sets of points x where $\rho(x, A) < \rho(x, B)$ or $\rho(x, A) > \rho(x, B)$.]
4. Let $X = \mathbf{R}$, and let $\mathscr{T}$ be the family of all sets of the form $U \cup (V \cap \mathbf{Q})$ where U, V are open in the usual sense. Then $\mathscr{T}$ is a topology which is

Hausdorff but not regular. (This shows, by the way, that a topology stronger than a normal topology need not be normal or even regular.)

5. Every separable metric space is second countable.

6. Let $\mathscr{E} = \{(a, b]: -\infty < a < b < \infty\}$.

(a) $\mathscr{E}$ is a base for a topology $\mathscr{T}$ on $\mathbf{R}$ in which the members of $\mathscr{E}$ are both open and closed.

(b) $\mathscr{T}$ is first countable but not second countable. (If $x \in \mathbf{R}$, every neighborhood base at x contains a set whose supremum is x.)

(c) $\mathbf{Q}$ is dense in $\mathbf{R}$ with respect to $\mathscr{T}$. [This shows that the converse of Proposition (4.5) is false.]

7. If X is a topological space, $x \in X$ is called a **cluster point** of the sequence $\{x_j\}$ if for every neighborhood U of x, $x_j \in U$ for infinitely many j. If X is first countable, x is a cluster point of $\{x_j\}$ iff some subsequence of $\{x_j\}$ converges to x.

8. If X is an infinite set with the cofinite topology and $\{x_j\}$ is a sequence of distinct points in X, then $x_j \to x$ for every $x \in X$.

9. If X is a linearly ordered set, the topology $\mathscr{T}$ generated by the sets $\{x: x < a\}$ and $\{x: x > a\}$ $(a \in X)$ is called the **order topology**.

(a) If $a, b \in X$ and $a < b$, there exist $U, V \in \mathscr{T}$ with $x \in U$, $b \in V$, and $x < y$ for all $x \in U$, $y \in V$. The order topology is the weakest topology with this property.

(b) If $Y \subset X$, the order topology on Y is never stronger than, but may be weaker than, the relative topology on Y induced by the order topology on X.

(c) The order topology on $\mathbf{R}$ is the usual topology.

10. A topological space X is called **disconnected** if there exist nonempty open sets U, V such that $U \cap V = \varnothing$ and $U \cup V = X$; otherwise X is **connected**. When we speak of connected or disconnected subsets of X, we refer to the relative topology on them.

(a) X is connected iff $\varnothing$ and X are the only subsets of X which are both open and closed.

(b) If $\{E_\alpha\}_{\alpha \in A}$ is a collection of connected subsets of X such that $\bigcap_{\alpha \in A} E_\alpha \neq \varnothing$, then $\bigcup_{\alpha \in A} E_\alpha$ is connected.

(c) If $A \subset X$ is connected, then $\overline{A}$ is connected.

(d) Every point $x \in X$ is contained in a maximal connected subset of X, and this subset is closed. (It is called the **connected component** of x.)

11. If $E_1, \ldots, E_n$ are subsets of a topological space, the closure of $\bigcup_1^n E_j$ is $\bigcup_1^n \overline{E}_j$.

12. Let X be a set. A **Kuratowski closure operator** on X is a map $A \to A^*$ from $\mathscr{P}(X)$ to $\mathscr{P}(X)$ satisfying (i) $\varnothing^* = \varnothing$, (ii) $A \subset A^*$ for all A, (iii) $(A^*)^* = A^*$ for all A, and (iv) $(A \cup B)^* = A^* \cup B^*$ for all A, B.

(a) If X is a topological space, the map $A \to \overline{A}$ is a Kuratowski closure operator. (Cf. Exercise 11.)

(b) Conversely, given a Kuratowski closure operator, let $\mathscr{F} = \{A \subset X: A = A^*\}$ and $\mathscr{T} = \{U \subset X: U^c \in \mathscr{F}\}$. Then $\mathscr{T}$ is a topology, and for any $A \subset X$, A^* is its closure with respect to $\mathscr{T}$.

13. If X is a topological space, U is open in X, and A is dense in X, then $\overline{U} = \overline{U \cap A}$.

2. CONTINUOUS MAPS

Topological spaces are the natural setting for the concept of continuity, which can be described in either global or local terms, as follows. Let X and Y be topological spaces. A map $f: X \to Y$ is called **continuous** if $f^{-1}(V)$ is open in X for every open $V \subset Y$, or equivalently, if $f^{-1}(K)$ is closed in X for every closed $K \subset Y$. If $x \in X$, f is called **continuous at** x if for every neighborhood V of $f(x)$ there is a neighborhood U of x such that $f(U) \subset V$, or equivalently, if $f^{-1}(V)$ is a neighborhood of x for every neighborhood V of $f(x)$. Clearly, if $f: X \to Y$ and $g: Y \to Z$ are continuous [or f is continuous at x and g is continuous at $f(x)$], then $g \circ f$ is continuous (at x). We shall denote the set of continuous maps from X to Y by $C(X, Y)$.

(4.8) Proposition. $f: X \to Y$ is continuous iff f is continuous at every $x \in X$.

Proof. If f is continuous and V is a neighborhood of $f(x)$, $f^{-1}(V^o)$ is a neighborhood of x, so f is continuous at x. Conversely, if f is continuous at each $x \in X$ and $V \subset Y$ is open, V is a neighborhood of each of its points, so $f^{-1}(V)$ is a neighborhood of each of its points. Thus $f^{-1}(V)$ is open, so f is continuous. □

(4.9) Proposition. If the topology on Y is generated by a family of sets $\mathscr{E}$, then $f: X \to Y$ is continuous iff $f^{-1}(V)$ is open in X for every $V \in \mathscr{E}$.

Proof. This is clear from Proposition (4.4) and the fact that f^{-1} commutes with unions and intersections. □

If $f: X \to Y$ is bijective and f and f^{-1} are both continuous, f is called a **homeomorphism,** and X and Y are said to be **homeomorphic**. In this case f^{-1} is a bijection from the open sets in Y to the open sets in X, so X and Y may be considered identical as far as their topological properties go. If $f: X \to Y$ is injective but not surjective, and $f: X \to \text{range}(f)$ is a homeomorphism when range(f) is given the relative topology, f is called a **homeomorphism onto its range.**

If X is any set and $\{f_\alpha\colon X \to Y_\alpha\}_{\alpha\in A}$ is a family of maps from X into some topological spaces Y_α, there is a unique weakest topology $\mathscr{T}$ on X which makes all the f_α's continuous. Namely, $\mathscr{T}$ is the topology generated by all sets of the form $f_\alpha^{-1}(U_\alpha)$ where U_α is open in Y_α and $\alpha \in A$. $\mathscr{T}$ is called the **weak topology generated by** $\{f_\alpha\}_{\alpha\in A}$. The most important example of this construction is the Cartesian product of topological spaces. If $\{X_\alpha\}_{\alpha\in A}$ is any family of topological spaces, the **product topology** on $X = \prod_{\alpha\in A} X_\alpha$ is the weak topology generated by the projection maps $\pi_\alpha\colon X \to X_\alpha$. If the index set A is finite, say $A = \{1,\dots,n\}$, the sets of the form $\prod_1^n U_j$, where U_j is open in X_j, are a base for the product topology, since $\prod_1^n U_j = \bigcap_1^n \pi_j^{-1}(U_j)$. When we consider a Cartesian product of topological spaces, we *always* endow it with the product topology, unless we specify otherwise.

(4.10) Proposition. If X_α is Hausdorff for each $\alpha \in A$, then $X = \prod_{\alpha\in A} X_\alpha$ is Hausdorff.

Proof. If x and y are distinct points of X, we have $\pi_\alpha(x) \neq \pi_\alpha(y)$ for some α. Let U and V be disjoint neighborhoods of $\pi_\alpha(x)$ and $\pi_\alpha(y)$ in X_α. Then $\pi_\alpha^{-1}(U)$ and $\pi_\alpha^{-1}(V)$ are disjoint neighborhoods of x and y in X. □

(4.11) Proposition. If X_α $(\alpha \in A)$ and Y are topological spaces and $X = \prod_{\alpha\in A} X_\alpha$, then $f\colon Y \to X$ is continuous iff $\pi_\alpha \circ f$ is continuous for each α.

Proof. If $\pi_\alpha \circ f$ is continuous for each α, then $f^{-1}(\pi_\alpha^{-1}(U_\alpha))$ is open in Y for all open U_α in X_α. By Proposition (4.9), f is continuous. The converse is obvious. □

If the spaces X_α are all equal to some fixed space X, the product $\prod_{\alpha\in A} X_\alpha$ is just the set X^A of all mappings from A to X, and the product topology is the topology of pointwise convergence. More precisely:

(4.12) Proposition. If X is a topological space, A is a set, and $\{f_n\}$ is a sequence in X^A, then $f_n \to f$ in the product topology iff $f_n \to f$ pointwise.

Proof. If $f_n \to f$ in the product topology, then $f_n(\alpha) = \pi_\alpha(f_n) \to \pi_\alpha(f) = f(\alpha)$ for all $\alpha \in A$. On the other hand, the sets

$$N(\alpha_1,\dots,\alpha_k,U_1,\dots,U_k) = \left\{g \in X^A\colon g(\alpha_j) \in U_j \text{ for } 1 \le j \le k\right\},$$

where $\alpha_j \in A$ and U_j is a neighborhood of $f(\alpha_j)$, form a neighborhood base for the product topology at f. If $f_n(\alpha_j) \to f(\alpha_j)$ for $j = 1,\dots,k$, then clearly $f_n \in N(\alpha_1,\dots,\alpha_k,U_1,\dots,U_k)$ for n sufficiently large; hence if $f_n \to f$ pointwise, then $f_n \to f$ in the product topology. □

We shall be particularly concerned with real- and complex-valued functions on topological spaces. If X is any set, we denote by $B(X,\mathbf{R})$ [resp. $B(X,\mathbf{C})$]

the space of all bounded real- (resp. complex-) valued functions on X. If X is a topological space, we also have the spaces $C(X,\mathbf{R})$ and $C(X,\mathbf{C})$ of continuous functions on X, and we define

$$BC(X,\mathbf{R}) = B(X,\mathbf{R}) \cap C(X,\mathbf{R})$$

and likewise $BC(X,\mathbf{C})$. In speaking of complex-valued functions we shall usually omit the $\mathbf{C}$ and write simply $B(X)$, $C(X)$, and $BC(X)$. Since addition and multiplication are continuous from $\mathbf{C}\times\mathbf{C}$ to $\mathbf{C}$, $C(X)$ and $BC(X)$ are complex vector spaces.

If $f \in B(X)$, we define the **uniform norm** of f to be

$$\|f\|_u = \sup\{|f(x)|\colon x \in X\}.$$

$B(X)$ thus becomes a metric space with the uniform metric $\rho(f,g) = \|f-g\|_u$. $B(X)$ is obviously complete: if $\{f_n\}$ is uniformly Cauchy, then $\{f_n(x)\}$ is Cauchy for each x, and if $f(x) = \lim f_n(x)$, it is easily verified that $\|f_n - f\|_u \to 0$.

(4.13) Proposition. If X is a topological space, $BC(X)$ is a closed subspace of $B(X)$ in the uniform metric; in particular, $BC(X)$ is complete.

Proof. Suppose $\{f_n\} \subset BC(X)$ and $\|f_n - f\|_u \to 0$. Given $\varepsilon > 0$, choose N so large that $\|f_n - f\| < \varepsilon/3$ for $n > N$. If $n > N$ and $x \in X$, since f_n is continuous at x there is a neighborhood U of x such that $|f_n(y) - f_n(x)| < \varepsilon/3$ for $y \in U$. But then

$$|f(y) - f(x)| \le |f(y) - f_n(y)| + |f_n(y) - f_n(x)| + |f_n(x) - f(x)| < \varepsilon,$$

so f is continuous at x. By Proposition (4.8), f is continuous. □

For a given topological space X it may happen that $C(X)$ consists only of constant functions. This is obviously the case, for example, if X has the trivial topology, but it can happen even when X is regular. Normal spaces, however, always have plenty of continuous functions, as the following fundamental theorems show.

(4.14) Lemma. Let X be a normal space. If A and B are disjoint closed sets in X, for each dyadic rational $r = k2^{-n} \in (0,1]$ there is an open set U_r such that $A \subset U_r \subset B^c$ and $\overline{U}_r \subset U_s$ for $r < s$.

Proof. Let $U_1 = B^c$. By normality, there exist disjoint open sets V, W such that $A \subset V$, $B \subset W$. Let $U_{1/2} = V$. Then W^c is closed, so we have

$$A \subset U_{1/2} \subset \overline{U}_{1/2} \subset W^c \subset B^c = U_1.$$

We now select U_r for $r = k2^{-n}$ by induction on n. Suppose we have chosen U_r for $r = k2^{-n}$, $0 < k < 2^n$, $1 \le n \le N-1$. To find U_r for $r = (2j+1)2^{-N}$ $(0 \le j < 2^{N-1})$, observe that $\overline{U}_{j2^{1-N}}$ and $(U_{(j+1)2^{1-N}})^c$ are disjoint closed sets (where, for $j = 0$, we set $\overline{U}_0 = A$), so as above we can choose an open U_r with

$$\overline{U}_{j2^{1-N}} \subset U_r \subset \overline{U}_r \subset U_{(j+1)2^{1-N}}.$$

These U_r's clearly have the desired properties. □

(4.15) Urysohn's Lemma. Let X be a normal space. If A and B are disjoint closed sets in X, there exists $f \in C(X,[0,1])$ such that $f = 0$ on A and $f = 1$ on B.

Proof. Let U_r be as in Lemma (4.14), and set $U_r = X$ for $r > 1$. Define $f(x) = \inf\{r\colon\ x \in U_r\}$. Since $A \subset U_r \subset B^c$ for $0 < r < 1$, we clearly have $f(x) = 0$ for $x \in A$, $f(x) = 1$ for $x \in B$, and $0 \le f(x) \le 1$ for all $x \in X$. We observe that $f(x) < \alpha$ iff $x \in U_r$ for some $r < \alpha$ iff $x \in \bigcup_{r<\alpha} U_r$, so $f^{-1}((-\infty,\alpha)) = \bigcup_{r<\alpha} U_r$ is open. Also, $f(x) > \alpha$ iff $x \notin U_r$ for some $r > \alpha$ iff $x \notin \overline{U}_s$ for some $s > \alpha$ (since $\overline{U}_s \subset U_r$ for $s < r$) iff $x \in \bigcup_{s>\alpha} (\overline{U}_s)^c$, so $f^{-1}((\alpha,\infty)) = \bigcup_{s>\alpha} (\overline{U}_s)^c$ is open. Since the open half-lines generate the topology on $\mathbf{R}$, f is continuous by Proposition (4.9). □

(4.16) The Tietze Extension Theorem. Let X be a normal space. If A is a closed subset of X and $f \in C(A,[a,b])$, there exists $F \in C(X,[a,b])$ such that $F|A = f$.

Proof. Replacing f by $(f-a)/(b-a)$, we may assume that $[a,b] = [0,1]$. We claim that there is a sequence $\{g_n\}$ of continuous functions on X such that $0 \le g_n \le 2^{n-1}/3^n$ on X and $0 \le f - \sum_1^n g_j \le (2/3)^n$ on A. To begin with, let $B = f^{-1}([0,1/3])$ and $C = f^{-1}([2/3,1])$. These are closed subsets of A, and since A itself is closed, they are closed in X. By Urysohn's lemma there is a continuous $g_1\colon X \to [0,1/3]$ with $g_1 = 0$ on B, $g_1 = 1/3$ on C: thus $0 \le f - g_1 \le 2/3$ on A. Having found $g_1,\dots,g_{n-1}$, by the same reasoning we can find $g_n\colon X \to [0, 2^{n-1}/3^n]$ such that $g_n = 0$ on the set where $f - \sum_1^{n-1} g_j \le 2^{n-1}/3^n$ and $g_n = 2^{n-1}/3^n$ on the set where $f - \sum_1^{n-1} g_j \ge (2/3)^n$. Let $F = \sum_1^\infty g_n$. Since $\|g_n\|_u \le 2^{n-1}/3^n$, the partial sums of this series converge uniformly, so F is continuous by Proposition (4.13). Moreover, on A we have $0 \le f - F \le (2/3)^n$ for all n, whence $F = f$ on A. □

(4.17) Corollary. If X is normal, $A \subset X$ is closed, and $f \in C(A)$, there exists $F \in C(X)$ such that $F|A = f$.

Proof. It suffices to assume that f is real-valued. Let $g = f/(1+|f|)$. Then $g \in C(A,(-1,1))$, so there exists $G \in C(X,[-1,1])$ with $G|A = g$. Let $B = G^{-1}(\{-1,1\})$. By Urysohn's lemma there exists $h \in C(X,[0,1])$ with $h = 1$ on A, $h = 0$ on B. Then $hG = G$ on A and $|hG| < 1$ everywhere, so $F = hG/(1-|hG|)$ does the job. □

A topological space X is called **completely regular** if X is T_1 and for each closed $A \subset X$ and $x \notin A$ there exists $f \in C(X,[0,1])$ such that $f(x) = 1$ and $f = 0$ on A. Completely regular spaces are also called **Tychonoff** spaces or $T_{3\frac{1}{2}}$ spaces. The latter terminology is justified, for every completely regular space is T_3 [if A, x, f are as above, then $f^{-1}((\frac{1}{2},\infty))$ and $f^{-1}((-\infty,\frac{1}{2}))$ are disjoint neighborhoods of x and A], and Urysohn's lemma shows that every T_4 space is completely regular.

EXERCISES

14. If X and Y are topological spaces, $f\colon X \to Y$ is continuous iff $f(\overline{A}) \subset \overline{f(A)}$ for all $A \subset X$ iff $\overline{f^{-1}(B)} \subset f^{-1}(\overline{B})$ for all $B \subset Y$.

15. If X is a topological space, $A \subset X$ is closed, and $g \in C(A)$ satisfies $g = 0$ on ∂A, then the extension of g to X defined by $g(x) = 0$ for $x \in A^c$ is continuous.

16. Let X be a topological space, Y a Hausdorff space, and f, g continuous maps from X to Y.

(a) $\{x\colon f(x) = g(x)\}$ is closed.

(b) If $f = g$ on a dense subset of X, then $f = g$ on all of X.

17. If X is a set, $\mathscr{F}$ a collection of real-valued functions on X, and $\mathscr{T}$ the weak topology generated by $\mathscr{F}$, then $\mathscr{T}$ is Hausdorff iff for every $x, y \in X$ with $x \ne y$ there exists $f \in \mathscr{F}$ with $f(x) \ne f(y)$.

18. Let $0'$ denote a point which is not an element of $(-1,1)$, and let $X = (-1,1) \cup \{0'\}$. Let $\mathscr{T}$ be the topology on X generated by the sets $(-1,a)$, $(a,1)$, $((-1,b)\setminus\{0\}) \cup \{0'\}$, and $((c,1)\setminus\{0\}) \cup \{0'\}$ where $-1 < a < 1$, $0 < b < 1$, and $-1 < c < 0$. [One should picture X as $(-1,1)$ with the point 0 split in two.] Then X is T_1 but not Hausdorff, although each point of X has a neighborhood which is homeomorphic to $(-1,1)$ (and hence is Hausdorff).

19. If X and Y are topological spaces and $y_0 \in Y$, then X is homeomorphic to $X \times \{y_0\}$ where the latter has the relative topology as a subset of $X \times Y$.

20. If $\{X_\alpha\}$ is a family of topological spaces, the space $X = \prod_\alpha X_\alpha$ (with the product topology) is uniquely determined up to homeomorphism by the following property: There exist continuous maps $\pi_\alpha\colon X \to X_\alpha$ such that if Y is any topological space and $f_\alpha \in C(Y, X_\alpha)$ for each α, there is a unique $F \in C(Y, X)$ such that $f_\alpha = \pi_\alpha \circ F$. (Thus X is the category-theoretic product of the X_α's in the category of topological spaces.)

21. If A is a countable set and X_α is a first (resp. second) countable space for each $\alpha \in A$, then $\prod_{\alpha\in A} X_\alpha$ is first (resp. second) countable.

22. If X is an infinite set with the cofinite topology, then every $f \in C(X)$ is constant.

23. Let X be a topological space, (Y, ρ) a complete metric space, and $\{f_n\}$ a sequence in Y^X such that $\sup_{x \in X} \rho(f_n(x), f_m(x)) \to 0$ as $m, n \to \infty$. There is a unique $f \in Y^X$ such that $\sup_{x \in X} \rho(f_n(x), f(x)) \to 0$ as $n \to \infty$. If each f_n is continuous, so is f.

24. Give an elementary proof of the Tietze extension theorem for the case $X = \mathbf{R}$.

25. A topological space X is normal iff X satisfies the conclusion of Urysohn's lemma iff X satisfies the conclusion of the Tietze extension theorem.

26. If $(X, \mathscr{T})$ is completely regular, then $\mathscr{T}$ is the weak topology generated by $C(X)$.

27. Let X and Y be topological spaces.

(a) If X is connected (cf. Exercise 10) and $f \in C(X, Y)$, then $f(X)$ is connected.

(b) X is called **arcwise connected** if for all $x, y \in X$ there exists $f \in C([0,1], X)$ with $f(0) = x$, $f(1) = y$. Every arcwise connected space is connected.

(c) Let $X = \{(x, y) \in \mathbf{R}^2\colon\ x > 0 \text{ and } y = \sin(x^{-1})\} \cup \{(0,0)\}$, with the relative topology induced from $\mathbf{R}^2$. Then X is connected but not arcwise connected.

28. If X_α is connected for each $\alpha \in A$, then $X = \prod_{\alpha \in A} X_\alpha$ is connected. [Fix $x \in X$ and let Y be the connected component of x in X. Show that Y includes $\{y \in X\colon \pi_\alpha(y) = \pi_\alpha(x) \text{ for all but finitely many } \alpha\}$ and that the latter set is dense in X. Use Exercises 10 and 19.]

29. Let X be a topological space equipped with an equivalence relation, $\tilde{X}$ the set of equivalence classes, and $\pi\colon X \to \tilde{X}$ the canonical projection. Then $\mathscr{T} = \{U \subset \tilde{X}\colon \pi^{-1}(U) \text{ is open in } X\}$ is a topology on $\tilde{X}$, called the **quotient topology**. If Y is a topological space, $f\colon \tilde{X} \to Y$ is continuous iff $f \circ \pi$ is continuous.

30. If X is a topological space and G is a group of homeomorphisms from X to itself, G induces an equivalence relation on X, namely, $x \sim y$ iff $y = g(x)$ for some $g \in G$. Let $X = \mathbf{R}^2$; describe the quotient space $\tilde{X}$ and the quotient topology on it (as in Exercise 29) for each of the following groups of linear maps. [For example, in (a), $\tilde{X}$ is homeomorphic to $[0, \infty)$.] Which separation axioms do these quotient topologies satisfy?

(a) $\left\{\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\colon \theta \in \mathbf{R}\right\}$.

(b) $\left\{\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}\colon a \in \mathbf{R}\right\}$.

(c) $\left\{\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}\colon a > 0,\ b \in \mathbf{R}\right\}$.

(d) $\left\{\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}\colon a \in \mathbf{Q}\right\}$.

3. NETS

As we have hinted above, sequential convergence does not play the same central role in general topological spaces as it does in metric spaces. The reasons for this may be illustrated by the following example. Consider the space $\mathbf{C}^{\mathbf{R}}$ of all complex-valued functions on $\mathbf{R}$, with the product topology, that is, the topology of pointwise convergence. $C(\mathbf{R})$ is a subspace of $\mathbf{C}^{\mathbf{R}}$. On the one hand, by Corollary (2.9), if $\{f_n\} \subset C(\mathbf{R})$ and $f_n \to f$ pointwise, then f is Borel measurable, so the set of limits of convergent sequences in $C(\mathbf{R})$ is a proper subset of $\mathbf{C}^{\mathbf{R}}$. Nonetheless, $C(\mathbf{R})$ is dense in $\mathbf{C}^{\mathbf{R}}$. Indeed, if $f \in \mathbf{C}^{\mathbf{R}}$, the sets

$$\{g \in \mathbf{C}^{\mathbf{R}}\colon |g(x_j) - f(x_j)| < \varepsilon \text{ for } j = 1, \ldots, n\}$$
$$(x_1, \ldots, x_n \in \mathbf{R},\ n \in \mathbf{N},\ \varepsilon > 0)$$

form a neighborhood base at f, and each of these sets clearly contains continuous functions.

There is, however, a generalization of the notion of sequence that works well in arbitrary topological spaces. A **directed set** is a set A equipped with a binary relation $\lesssim$ such that (i) $\alpha \lesssim \alpha$ for all $\alpha \in A$; (ii) if $\alpha \lesssim \beta$ and $\beta \lesssim \gamma$ then $\alpha \lesssim \gamma$, and (iii) for any $\alpha, \beta \in A$ there exists $\gamma \in A$ such that $\alpha \lesssim \gamma$ and $\beta \lesssim \gamma$. (If $\alpha \lesssim \beta$ we shall also write $\beta \gtrsim \alpha$.) A **net** in a set X is a mapping $\alpha \to x_\alpha$ from a directed set A into X. We shall usually denote such a mapping by $\langle x_\alpha \rangle_{\alpha \in A}$, and we say that $\langle x_\alpha \rangle$ is indexed by A.

Let us see some examples of directed sets.

(i) The set of positive integers $\mathbf{N}$, with $j \lesssim k$ iff $j \leq k$.

(ii) The set $\mathbf{R} \setminus \{a\}$ $(a \in \mathbf{R})$, with $x \lesssim y$ iff $|x - a| \geq |y - a|$.

(iii) The set of all partitions $a = x_0 < \cdots < x_n = b$ of the interval $[a, b]$, with $(a = x_0 < \cdots < x_n = b) \lesssim (a = y_0 < \cdots < y_m = b)$ iff $\max(x_j - x_{j-1}) \geq \max(y_k - y_{k-1})$.

(iv) The set $\mathscr{N}$ of all neighborhoods of a point x in a topological space X, with $U \lesssim V$ iff $U \supset V$. We say that $\mathscr{N}$ is directed by **reverse inclusion.**

(v) If A and B are directed sets, there is a canonical way of making $A \times B$ into a directed set: namely, $(\alpha, \beta) \lesssim (\alpha', \beta')$ iff $\alpha \lesssim \alpha'$ and $\beta \lesssim \beta'$.

Examples (i)–(iii) are familiar from elementary analysis: a net indexed by $\mathbf{N}$ is just a sequence, and nets indexed by the sets in (ii) and (iii) occur in studying

limits of real variables and Riemann integrals. Example (iv) is of fundamental importance in topology, and we shall see several uses of the construction in (v).

Let X be a topological space and E a subset of X. A net $\langle x_\alpha \rangle_{\alpha \in A}$ is **eventually** in E if there exists $\alpha_0 \in A$ such that $x_\alpha \in E$ for $\alpha \gtrsim \alpha_0$, and $\langle x_\alpha \rangle$ is **frequently** in E if for every $\alpha \in A$ there exists $\beta \gtrsim \alpha$ such that $x_\beta \in E$. A point $x \in X$ is a **limit** of $\langle x_\alpha \rangle$ (or $\langle x_\alpha \rangle$ **converges** to x, or $x_\alpha \to x$) if for every neighborhood U of x, $\langle x_\alpha \rangle$ is eventually in U, and x is a **cluster point** of $\langle x_\alpha \rangle$ if for every neighborhood U of x, $\langle x_\alpha \rangle$ is frequently in U. The next three propositions show that nets are a good substitute for sequences.

(4.18) Proposition. If X is a topological space, $E \subset X$, and $x \in X$, then x is an accumulation point of E iff there is a net in $E \setminus \{x\}$ which converges to x, and $x \in \bar{E}$ iff there is a net in E which converges to x.

Proof. If x is an accumulation point of E, let $\mathcal{N}$ be the set of neighborhoods of x, directed by reverse inclusion. For each $U \in \mathcal{N}$ pick $x_U \in (U \setminus \{x\}) \cap E$. Then $x_U \to x$. Conversely, if $x_\alpha \in E \setminus \{x\}$ and $x_\alpha \to x$, then every deleted neighborhood of x contains some x_α, so x is an accumulation point of E. Likewise, if $x_\alpha \to x$ where $x_\alpha \in E$, then $x \in \bar{E}$, and the converse follows from Proposition (4.1). □

(4.19) Proposition. If X and Y are topological spaces and $f\colon X \to Y$, then f is continuous at $x \in X$ iff for every net $\langle x_\alpha \rangle$ converging to x, $\langle f(x_\alpha) \rangle$ converges to $f(x)$.

Proof. If f is continuous at x and V is a neighborhood of $f(x)$, then $f^{-1}(V)$ is a neighborhood of x. Hence, if $x_\alpha \to x$ then $\langle x_\alpha \rangle$ is eventually in $f^{-1}(V)$, so $\langle f(x_\alpha) \rangle$ is eventually in V, and thus $f(x_\alpha) \to f(x)$. Conversely, if f is not continuous at x, there is a neighborhood V of $f(x)$ such that $f^{-1}(V)$ is not a neighborhood of x; that is, $x \notin (f^{-1}(V))^o$, or equivalently $x \in \overline{f^{-1}(V^c)}$. By Proposition (4.18), there is a net $\langle x_\alpha \rangle$ in $f^{-1}(V^c)$ which converges to x. But then $f(x_\alpha) \notin V$, so $f(x_\alpha) \not\to f(x)$. □

A **subnet** of a net $\langle x_\alpha \rangle_{\alpha \in A}$ is a net $\langle y_\beta \rangle_{\beta \in B}$ together with a map $\beta \to \alpha_\beta$ from B to A such that (i) for every $\alpha_0 \in A$ there exists $\beta_0 \in B$ such that if $\beta \gtrsim \beta_0$ then $\alpha_\beta \gtrsim \alpha_0$, and (ii) $y_\beta = x_{\alpha_\beta}$. Warning: the name "subnet" is used because subnets perform much the same functions as subsequences, but it should not be taken too literally, as the index set B may well have larger cardinality than the index set A. In particular, a subnet of a sequence need not be a subsequence. (This is true even if $B = \mathbf{N}$: see Exercise 32.)

(4.20) Proposition. If $\langle x_\alpha \rangle_{\alpha \in A}$ is a net in a topological space X, then $x \in X$ is a cluster point of $\langle x_\alpha \rangle$ iff $\langle x_\alpha \rangle$ has a subnet which converges to x.

Proof. If $\langle y_\beta \rangle = \langle x_{\alpha_\beta} \rangle$ is a subnet converging to x and U is a neighborhood of x, choose $\beta_1 \in B$ such that $y_\beta \in U$ for $\beta \gtrsim \beta_1$. Also, given $\alpha \in A$,

choose $\beta_2 \in B$ such that $\alpha_\beta \gtrsim \alpha$ for $\beta \gtrsim \beta_2$. Then there exists $\beta \in B$ with $\beta \gtrsim \beta_1$ and $\beta \gtrsim \beta_2$, and we have $\alpha_\beta \gtrsim \alpha$ and $x_{\alpha_\beta} = y_\beta \in U$. Thus $\langle x_\alpha \rangle$ is frequently in U, so x is a cluster point of $\langle x_\alpha \rangle$. Conversely, if x is a cluster point of $\langle x_\alpha \rangle$, let $\mathcal{N}$ be the set of neighborhoods of x and make $\mathcal{N} \times A$ into a directed set by declaring that $(U, \alpha) \lesssim (U', \alpha')$ if $U \supset U'$ and $\alpha \lesssim \alpha'$. For each $(U, \gamma) \in \mathcal{N} \times A$ we can choose $\alpha_{(U,\gamma)} \in A$ such that $\alpha_{(U,\gamma)} \gtrsim \gamma$ and $x_{\alpha_{(U,\gamma)}} \in U$. Then if $(U', \gamma') \gtrsim (U, \gamma)$ we have $\alpha_{(U',\gamma')} \gtrsim \gamma' \gtrsim \gamma$ and $x_{\alpha_{(U',\gamma')}} \in U' \subset U$, whence it follows that $\langle x_{\alpha_{(U,\gamma)}} \rangle$ is a subnet of $\langle x_\alpha \rangle$ which converges to x. □

EXERCISES

31. If A is a directed set, a subset B of A is called **cofinal** in A if for each $\alpha \in A$ there exists $\beta \in B$ such that $\beta \gtrsim \alpha$.

(a) If B is cofinal in A and $\langle x_\alpha \rangle_{\alpha \in A}$ is a net, the inclusion map $B \to A$ makes $\langle x_\beta \rangle_{\beta \in B}$ a subnet of $\langle x_\alpha \rangle_{\alpha \in A}$.

(b) If $\langle x_\alpha \rangle_{\alpha \in A}$ is a net in a topological space, then x_α converges to x iff for every cofinal $B \subset A$ there is a cofinal $C \subset B$ such that $\langle x_\gamma \rangle_{\gamma \in C}$ converges to x.

32. Let $\langle x_n \rangle_{n \in \mathbf{N}}$ be a sequence.

(a) If $k \to n_k$ is a map from $\mathbf{N}$ to itself, $\langle x_{n_k} \rangle_{k \in \mathbf{N}}$ is a subnet of $\langle x_n \rangle$ iff $n_k \to \infty$ as $k \to \infty$, and it is a subsequence (as defined in the Prologue) iff n_k is strictly increasing in k.

(b) There is a natural one-to-one correspondence between the subsequences of $\langle x_n \rangle$ and the subnets of $\langle x_n \rangle$ defined by cofinal sets as in Exercise 31.

33. A topological space X is Hausdorff iff every net in X converges to at most one point. [If X is not Hausdorff, let x and y be distinct points with no disjoint neighborhoods, and let $\mathcal{N}_x$, $\mathcal{N}_y$ be the sets of neighborhoods of x and y. Consider the directed set $\mathcal{N}_x \times \mathcal{N}_y$, where $(U, V) \lesssim (U', V')$ iff $U \supset U'$ and $V \supset V'$.]

34. Let $\langle x_\alpha \rangle_{\alpha \in A}$ be a net in a topological space, and for each $\alpha \in A$ let $E_\alpha = \{x_\beta\colon \beta \gtrsim \alpha\}$. Then x is a cluster point of $\langle x_\alpha \rangle$ iff $x \in \bigcap_{\alpha \in A} \bar{E}_\alpha$.

35. If X has the weak topology generated by a family $\mathcal{F}$ of functions, then $\langle x_\alpha \rangle$ converges to $x \in X$ iff $\langle f(x_\alpha) \rangle$ converges to $f(x)$ for all $f \in \mathcal{F}$. [In particular, if $X = \prod_{\beta \in B} X_\beta$, then $x_\alpha \to x$ iff $\pi_\beta(x_\alpha) \to \pi_\beta(x)$ for all $\beta \in B$.]

36. Let X be a set and $\mathcal{A}$ the collection of finite subsets of X, directed by inclusion. Let $f\colon X \to \mathbf{R}$ be an arbitrary function, and for $A \in \mathcal{A}$, let $z_A = \sum_{x \in A} f(x)$. Then the net $\langle z_A \rangle$ converges in $\mathbf{R}$ iff $\{x\colon f(x) \neq 0\}$ is a countable set $\{x_n\}$ and $\sum_1^\infty |f(x_n)| < \infty$, in which case $z_A \to \sum_1^\infty f(x_n)$. [Cf. Proposition (P.20).]

37. Let X be the set of Lebesgue measurable complex-valued functions on $[0, 1]$. There is no topology $\mathcal{T}$ on X such that a sequence $\langle f_n \rangle$ converges to f with respect to $\mathcal{T}$ iff $f_n \to f$ a.e. (Use Corollary 2.32 and Exercise 31b.)

4. COMPACT SPACES

In the Prologue we gave three equivalent characterizations of compactness for metric spaces: the Heine–Borel property, the Bolzano–Weierstrass property, and the property of being complete and totally bounded. Only the first two of these make sense for general topological spaces, and it is the first one that turns out to be the most useful. Accordingly, we define a topological space X to be **compact** if whenever $\{U_\alpha\}_{\alpha \in A}$ is an open cover of X—that is, a collection of open sets such that $X = \bigcup_{\alpha \in A} U_\alpha$—there is a finite subset B of A such that $X = \bigcup_{\alpha \in B} U_\alpha$. To be brief (and with a disregard for elegance of diction, since the adjectives "open" and "finite" refer to different things), we say: X is compact if every open cover of X has a finite subcover. A subset Y of a topological space X is called compact if it is compact in the relative topology; thus $Y \subset X$ is compact iff whenever $\{U_\alpha\}_{\alpha \in A}$ is a collection of open subsets of X with $Y \subset \bigcup_{\alpha \in A} U_\alpha$, there is a finite $B \subset A$ with $Y \subset \bigcup_{\alpha \in B} U_\alpha$. Furthermore, Y is called **precompact** if its closure is compact.

DeMorgan's laws lead to the following characterization of compactness in terms of closed sets. A family $\{F_\alpha\}_{\alpha \in A}$ of subsets of X has the **finite intersection property** if $\bigcap_{\alpha \in B} F_\alpha \neq \varnothing$ for all finite $B \subset A$.

(4.21) Proposition. A topological space X is compact iff for every family $\{F_\alpha\}_{\alpha \in A}$ of closed sets with the finite intersection property, $\bigcap_{\alpha \in A} F_\alpha \neq \varnothing$.

Proof. Let $U_\alpha = (F_\alpha)^c$. Then U_α is open, $\bigcap_{\alpha \in A} F_\alpha \neq \varnothing$ iff $\bigcup_{\alpha \in A} U_\alpha \neq X$, and $\{F_\alpha\}$ has the finite intersection property iff no finite subfamily of $\{U_\alpha\}$ covers X. The result follows. □

We now list several basic facts about compact spaces.

(4.22) Proposition. A closed subset of a compact space is compact.

Proof. If X is compact, $F \subset X$ is closed, and $\{U_\alpha\}_{\alpha \in A}$ is a family of open subsets of X with $F \subset \bigcup_{\alpha \in A} U_\alpha$, then $\{U_\alpha\}_{\alpha \in A} \cup \{F^c\}$ is an open cover of X. It has a finite subcover, so by discarding F^c if necessary we obtain a finite subcollection of $\{U_\alpha\}_{\alpha \in A}$ which covers F. □

(4.23) Proposition. If F is a compact subset of a Hausdorff space X and $x \in F^c$, there are disjoint open sets U, V such that $x \in U$ and $F \subset V$.

Proof. For each $y \in F$ choose disjoint U_y, V_y with $x \in U_y$, $y \in V_y$. $\{V_y\}_{y \in F}$ is an open cover of F, so it has a finite subcover $\{V_{y_j}\}_1^n$. Then $U = \bigcap_1^n U_{y_j}$ and $V = \bigcup_1^n V_{y_j}$ do the job. □

(4.24) Proposition. Every compact subset of a Hausdorff space is closed.

Proof. According to Proposition (4.23), F^c is a neighborhood of each of its points, hence is open. □

(4.25) Proposition. Every compact Hausdorff space is normal.

Proof. Suppose X is compact Hausdorff and E, F are disjoint closed subsets of X. By Proposition (4.23), for each $x \in E$ there exist disjoint open U_x, V_x with $x \in U_x$, $F \subset V_x$. E is compact by Proposition (4.22) and $\{U_x\}_{x \in E}$ is an open cover of E, so it has a finite subcover $\{U_{x_j}\}_1^n$. Let $U = \bigcup_1^n U_{x_j}$, $V = \bigcap_1^n V_{x_j}$. Then U, V are disjoint open sets with $E \subset U$, $F \subset V$. □

(4.26) Proposition. If X is compact and $f: X \to Y$ is continuous, then $f(X)$ is compact.

Proof. Let $\{U_\alpha\}$ be an open cover of $f(X)$ in Y. Then $\{f^{-1}(U_\alpha)\}$ is an open cover of X, so it has a finite subcover $\{f^{-1}(U_{\alpha_j})\}$, and $\{U_{\alpha_j}\}$ is then a finite subcover of $f(X)$. □

(4.27) Corollary. If X is compact, then $C(X) = BC(X)$.

(4.28) Proposition. If X is compact and Y is Hausdorff, then any continuous bijection $f: X \to Y$ is a homeomorphism.

Proof. If $E \subset X$ is closed, then E is compact, hence $f(E)$ is compact, hence $f(E)$ is closed, by Propositions (4.22), (4.24), and (4.26). This means that f^{-1} is continuous, so f is a homeomorphism. □

We now show that a version of the Bolzano–Weierstrass property holds for compact topological spaces. As one might suspect, it is merely necessary to replace sequences by nets.

(4.29) Theorem. If X is a topological space, the following are equivalent:

(a) X is compact.

(b) Every net in X has a cluster point.

(c) Every net in X has a convergent subnet.

Proof. The equivalence of (b) and (c) follows from Proposition (4.20). If X is compact and $\langle x_\alpha \rangle$ is a net in X, let F_α be the closure of $\{x_\beta : \beta \gtrsim \alpha\}$. Clearly the family $\{F_\alpha\}$ has the finite intersection property, so by Proposition (4.21) there exist points in $\bigcap_\alpha F_\alpha$, and any such point is a cluster point of $\langle x_\alpha \rangle$ (Exercise 34). On the other hand, if X is not compact, let $\{U_\beta\}_{\beta \in B}$ be an open

cover of X with no finite subcover. Let $\mathscr{A}$ be the collection of finite subsets of B, directed by inclusion, and for each $A \in \mathscr{A}$ let x_A be a point in $(\bigcup_{\beta \in A} U_\beta)^c$. Then $\langle x_A \rangle_{A \in \mathscr{A}}$ is a net with no cluster point: indeed, if $x \in X$, choose $\beta \in B$ with $x \in U_\beta$. If $\beta \in A$, that is, if $A \geq \{\beta\}$, then $x_A \notin U_\beta$, so x is not a cluster point of $\langle x_A \rangle$. □

There are two other useful concepts related to compactness. A topological space X is called **countably compact** if every countable open cover of X has a finite subcover, and **sequentially compact** if every sequence in X has a convergent subsequence. Of course, every compact space is countably compact, and for metric spaces compactness and sequential compactness are equivalent. Other relationships are explored in the exercises.

EXERCISES

38. Let $(X, \mathscr{T})$ be a compact Hausdorff space. If $\mathscr{T}'$ is another topology on X which is strictly stronger than $\mathscr{T}$, then $\mathscr{T}'$ is Hausdorff but not compact. If $\mathscr{T}'$ is strictly weaker than $\mathscr{T}$, then $\mathscr{T}'$ is compact but not Hausdorff.

39. Every sequentially compact space is countably compact.

40. If X is countably compact, then every sequence in X has a cluster point. If in addition X is first countable, then X is sequentially compact.

41. The set of countable ordinals, with the order topology (Exercise 9), is sequentially compact and first countable but not compact. [To prove sequential compactness, use Proposition (P.19).]

42. If $x \in [0,1]$, let $\sum_1^\infty a_n(x)2^{-n}$ $[a_n(x) = 0$ or $1]$ be the base-2 decimal expansion of x. [If x is a dyadic rational, choose the expansion such that $a_n(x) = 0$ for n large.] Then the sequence $\langle a_n \rangle$ in $\{0,1\}^{[0,1]}$ has no pointwise convergent subsequence. (Hence $\{0,1\}^{[0,1]}$, with the product topology, is not sequentially compact. It is, however, compact, as we shall show in Section 6.)

43. Every compact metric space is separable. (Use total boundedness.)

44. If X is countably compact and $f\colon X \to Y$ is continuous, then $f(X)$ is countably compact.

45. If X is normal, then X is countably compact iff $C(X) = BC(X)$. [Use Exercises 40 and 44. If $\langle x_n \rangle$ is a sequence in X with no cluster point, then $\{x_n\colon n \in \mathbf{N}\}$ is closed, and Corollary (4.17) applies.]

5. LOCALLY COMPACT SPACES

A topological space is called **locally compact** if every point has a compact neighborhood. We shall be mainly concerned with locally compact Hausdorff spaces, which we call **LCH** spaces for short.

(4.30) Proposition. If X is an LCH space, $U \subset X$ is open, and $x \in U$, there is a compact neighborhood N of x such that $N \subset U$.

Proof. We may assume $\overline{U}$ is compact; otherwise, replace U by $U \cap F^o$ where F is a compact neighborhood of x. By Proposition (4.23) there are disjoint relatively open sets V, W in $\overline{U}$ with $x \in V$ and $\partial U \subset W$. Since $V \subset U$, V is open in X, and $\overline{V}$ is a closed and hence compact subset of $U \setminus W$. Thus we may take $N = \overline{V}$. □

(4.31) Proposition. If X is an LCH space and $K \subset U \subset X$ where K is compact and U is open, there exists a precompact open V such that $K \subset V \subset \overline{V} \subset U$.

Proof. By Proposition (4.30), for each $x \in K$ we can choose a compact neighborhood N_x of x with $N_x \subset U$. Then $\{N_x^o\}_{x \in K}$ is an open cover of K, so there is a finite subcover $\{N_{x_j}^o\}_1^n$. Let $V = \bigcup_1^n N_{x_j}^o$: then $K \subset V$ and $\overline{V} = \bigcup_1^n N_{x_j} \subset U$. □

(4.32) Urysohn's Lemma, Locally Compact Version. If X is an LCH space and $K \subset U \subset X$ where K is compact and U is open, there exists $f \in C(X, [0,1])$ such that $f = 1$ on K and $f = 0$ outside a compact subset of U.

Proof. Let V be as in Proposition (4.31). Then $\overline{V}$ is normal by Proposition (4.25), so by Urysohn's lemma (4.15) there exists $f \in C(\overline{V}, [0,1])$ such that $f = 1$ on K and $f = 0$ on ∂V. We extend f to X by setting $f = 0$ on $\overline{V}^c$. Then if $E \subset [0,1]$ is closed, $f^{-1}(E) = (f|\overline{V})^{-1}(E)$ if $0 \notin E$ and $f^{-1}(E) = (f|\overline{V})^{-1}(E) \cup V^c$ if $0 \in E$. These sets are closed, so f is continuous. □

(4.33) Corollary. Every LCH space is completely regular.

(4.34) Tietze Extension Theorem, Locally Compact Version. Suppose X is an LCH space and $K \subset X$ is compact. If $f \in C(K)$, there exists $F \in C(X)$ such that $F|K = f$. Moreover, F may be taken to vanish outside a compact set.

The proof is similar to that of Theorem (4.32); details are left to the reader (Exercise 46).

The preceding results show that LCH spaces have a rich supply of continuous functions which vanish outside compact sets. It is convenient to introduce some terminology: if X is a topological space and $f \in C(X)$, the **support** of f, denoted by $\operatorname{supp}(f)$, is the smallest closed set outside of which f vanishes, that is, the closure of $f^{-1}(\{0\}^c)$. If $\operatorname{supp}(f)$ is compact, we say that f is **compactly supported**, and we define

$$C_c(X) = \{f \in C(X)\colon \operatorname{supp}(f) \text{ is compact}\}.$$

Moreover, if $f \in C(X)$, we say that f **vanishes at** ∞ if for every $\varepsilon > 0$ the set

$\{x: |f(x)| \geq \varepsilon\}$ is compact, and we define

$$C_0(X) = \{f \in C(X): f \text{ vanishes at } \infty\}.$$

Clearly $C_c(X) \subset C_0(X) \subset BC(X)$.

(4.35) Proposition. If X is an LCH space, $C_0(X)$ is the closure of $C_c(X)$ in the uniform metric.

Proof. If $\{f_n\}$ is a sequence in $C_c(X)$ which converges to $f \in C(X)$ uniformly, for each $\varepsilon > 0$ there exists $n \in \mathbf{N}$ such that $\|f_n - f\|_u < \varepsilon$. Then $|f(x)| < \varepsilon$ if $x \notin \text{supp}(f_n)$, so $f \in C_0(X)$. Conversely, if $f \in C_0(X)$, for $n \in \mathbf{N}$ let $K_n = \{x: |f(x)| \geq n^{-1}\}$. Then K_n is compact, so by Theorem (4.32) there exists $g_n \in C_c(X)$ with $0 \leq g_n \leq 1$ and $g_n = 1$ on K_n. Let $f_n = g_n f$. Then $f_n \in C_c(X)$ and $\|f_n - f\|_u \leq n^{-1}$, so $f_n \to f$ uniformly. □

If X is a noncompact LCH space, it is possible to make X into a compact Hausdorff space by adding a single point "at infinity" in such a way that the functions in $C_0(X)$ are precisely those continuous functions f such that $f(x) \to 0$ as x approaches the point at infinity. More precisely, let ∞ denote a point (i.e., set) which is not an element of X, let $X^* = X \cup \{\infty\}$, and let $\mathscr{T}$ be the collection of all subsets U of X^* such that either (i) U is an open subset of X, or (ii) $\infty \in U$ and U^c is a compact subset of X.

(4.36) Proposition. If X, X^*, and $\mathscr{T}$ are as above, then $\mathscr{T}$ is a compact Hausdorff topology on X^*, and the inclusion map $i: X \to X^*$ is a homeomorphism onto its range. Moreover, if $f \in C(X)$, then f extends continuously to X^* iff $f = g + c$ where $g \in C_0(X)$ and c is a constant, in which case the continuous extension is given by $g(\infty) = c$.

The proof is straightforward and is left to the reader (Exercise 47). X^* is called the **one-point compactification** or **Alexandrov compactification** of X.

If X is a topological space, the space $\mathbf{C}^X$ of all complex-valued functions on X can be topologized in various ways. One way, of course, is the product topology, that is, the topology of pointwise convergence. Another is the **topology of uniform convergence,** which is generated by the sets

$$\{g \in \mathbf{C}^X: \sup_{x \in X} |g(x) - f(x)| < n^{-1}\} \quad (n \in \mathbf{N}, f \in \mathbf{C}^X).$$

The proof of Proposition (4.13) shows that $C(X)$ is a closed subspace of $\mathbf{C}^X$ in the topology of uniform convergence. Intermediate between these two topologies is the **topology of uniform convergence on compact sets,** which is generated by the sets

$$\{g \in \mathbf{C}^X: \sup_{x \in K} |g(x) - f(x)| < n^{-1}\} \quad (n \in \mathbf{N}, f \in \mathbf{C}^X, K \subset X \text{ compact}).$$

We shall now examine this topology in the case where X is an LCH space.

(4.37) Lemma. If X is an LCH space and $E \subset X$, then E is closed iff $E \cap K$ is closed for every compact $K \subset X$.

Proof. If E is not closed, pick $x \in \bar{E} \setminus E$ and let K be a compact neighborhood of x. Then x is an accumulation point of $K \cap E$ but is not in $K \cap E$, so by Proposition (4.1) $K \cap E$ is not closed. The converse is obvious. □

(4.38) Proposition. If X is an LCH space, $C(X)$ is a closed subspace of $\mathbf{C}^X$ in the topology of uniform convergence on compact sets.

Proof. If f is in the closure of $C(X)$, then f is a uniform limit of continuous functions on each compact $K \subset X$, so $f|K$ is continuous. If $E \subset \mathbf{C}$ is closed, $f^{-1}(E) \cap K = (f|K)^{-1}(E)$ is thus closed for each compact K, so by Lemma (4.37) $f^{-1}(E)$ is closed, whence f is continuous. □

A topological space X is called **σ-compact** if it is a countable union of compact sets. To appreciate the significance of the next two propositions, see Exercise 53.

(4.39) Proposition. If X is a σ-compact LCH space, there is a sequence $\{U_n\}$ of precompact open sets such that $X = \bigcup_1^\infty U_n$.

Proof. Suppose $X = \bigcup_1^\infty K_n$ where each K_n is compact, and for each $x \in X$ let V_x be a precompact open neighborhood of x. For each n there is a finite family of V_x's which cover K_n, and we can take U_n to be the union of this family. □

(4.40) Proposition. If X is a σ-compact LCH space and $\{U_n\}$ is as in Proposition (4.39), then for each $f \in \mathbf{C}^X$ the sets

$$\{g \in \mathbf{C}^X: \sup_{x \in \bar{U}_n} |g(x) - f(x)| < m^{-1}\} \quad (m, n \in \mathbf{N})$$

form a neighborhood base for f in the topology of uniform convergence on compact sets. Hence this topology is first countable, and $f_j \to f$ uniformly on compact sets iff $f_j \to f$ uniformly on each $\bar{U}_n$.

Proof. These assertions follow easily from the observation that every compact subset of X is contained in the union of finitely many $\bar{U}_n$'s. Details are left to the reader (Exercise 48). □

We close this section with a construction that is useful in a number of situations. If X is a topological space and $E \subset X$, a **partition of unity** on E is a collection $\{h_\alpha\}_{\alpha \in A}$ of functions in $C(X, [0,1])$ such that (i) each $x \in X$ has a neighborhood on which only finitely many h_α's are nonzero, and (ii) $\sum_{\alpha \in A} h_\alpha(x) = 1$ for $x \in E$. A partition of unity $\{h_\alpha\}$ is **subordinate** to an open cover $\mathscr{U}$ of E if for each α there exists $U \in \mathscr{U}$ with $\text{supp}(h_\alpha) \subset U$.

(4.41) Proposition. Let X be an LCH space, K a compact subset of X, and $\{U_j\}_1^n$ an open cover of K. There is a partition of unity on K subordinate to $\{U_j\}_1^n$ consisting of compactly supported functions.

Proof. By Proposition (4.30), each $x \in K$ has a compact neighborhood N_x such that $N_x \subset U_j$ for some j. Since $\{N_x^o\}$ is an open cover of K, there exist $x_1, \ldots, x_m$ such that $K \subset \bigcup_1^m N_{x_k}$. Let F_j be the union of those N_{x_k}'s which are subsets of U_j. By Urysohn's lemma there exist $g_1, \ldots, g_n \in C_c(X)$ with $g_j = 1$ on F_j and $\operatorname{supp}(g_j) \subset U_j$. Then $\Sigma_1^n g_k \geq 1$ on K, so by Urysohn again there exists $f \in C_c(X)$ with $f = 1$ on K and $\operatorname{supp}(f) \subset \{x: \Sigma_1^n g_k(x) > 0\}$. Let $g_{n+1} = 1 - f$, so that $\Sigma_1^{n+1} g_k > 0$ everywhere, and for $j = 1, \ldots, n$ let $h_j = g_j / \Sigma_1^{n+1} g_k$. Then $\operatorname{supp}(h_j) = \operatorname{supp}(g_j) \subset U_j$ and $\Sigma_1^n h_j = 1$ on K. □

A generalization of this result may be found in Exercise 56.

EXERCISES

46. Prove Theorem (4.34).

47. Prove Proposition (4.36). Also, show that if X is Hausdorff but not locally compact, Proposition (4.36) remains valid except that X^* is not Hausdorff.

48. Complete the proof of Proposition (4.40).

49. Let X be a compact Hausdorff space.

(a) If $E \subset X$ is open, then E is locally compact in the relative topology.

(b) If E is dense in X and locally compact in the relative topology, then E is open in X. (Use Exercise 13.)

50. Let U be an open subset of a compact Hausdorff space X and U^* its one-point compactification (cf. Exercise 49a). If φ: $X \to U^*$ is defined by $\varphi(x) = x$ if $x \in U$, $\varphi(X) = \infty$ if $x \in U^c$, then φ is continuous.

51. If X, Y are topological spaces, $\varphi \in C(X, Y)$ is called **proper** if $\varphi^{-1}(K)$ is compact in X whenever K is compact in Y. Suppose X, Y are LCH spaces and X^*, Y^* are their one-point compactifications. If $\varphi \in C(X, Y)$, define φ^*: $X^* \to Y^*$ by $\varphi^*(x) = \varphi(x)$ for $x \in X$ and $\varphi^*(\infty_X) = \infty_Y$. Then φ^* is continuous iff φ is proper.

52. Lemma (4.37) remains true if the assumption that X is locally compact is replaced by the assumption that X is first countable.

53. Let $\mathbf{Q}$ have the relative topology induced from $\mathbf{R}$.

(a) $\mathbf{Q}$ is not locally compact.

(b) $\mathbf{Q}$ is σ-compact (it is a countable union of singleton sets), but uniform convergence on singletons (i.e., pointwise convergence) does not imply uniform convergence on compact subsets of $\mathbf{Q}$.

54. Every open set in a second countable LCH space is σ-compact.

55. (a) If (Y, ρ) is a metric space then

$$\rho^*(y, z) = \frac{\rho(y, z)}{1 + \rho(y, z)}$$

is a bounded metric on Y which defines the same topology as ρ. [Hint: $1/(1 + t)$ is increasing in t.]

(b) If X is a topological space, the topology on $\mathbf{C}^X$ of uniform convergence is induced by the metric

$$\rho(f, g) = \sup_{x \in X} \frac{|f(x) - g(x)|}{1 + |f(x) - g(x)|}.$$

(c) If X is a σ-compact LCH space and $\{U_n\}_1^\infty$ is as in Proposition (4.39), the topology on $\mathbf{C}^X$ of uniform convergence on compact sets is induced by the metric

$$\rho(f, g) = \sum_1^\infty \sup_{x \in \bar{U}_n} \frac{2^{-n}|f(x) - g(x)|}{1 + |f(x) - g(x)|}.$$

(d) Generalize (b) and (c), as well as the results in the text, to the space Y^X where Y is an arbitrary metric space.

56. An open cover $\mathscr{U}$ of a topological space X is called **locally finite** if each $x \in X$ has a neighborhood which intersects only finitely many members of $\mathscr{U}$. If $\mathscr{U}, \mathscr{V}$ are open covers of X, $\mathscr{V}$ is a **refinement** of $\mathscr{U}$ if for each $V \in \mathscr{V}$ there exists $U \in \mathscr{U}$ with $V \subset U$. X is called **paracompact** if every open cover of X has a locally finite refinement.

(a) If X is a σ-compact LCH space, then X is paracompact. In fact, every open cover $\mathscr{U}$ has locally finite refinements $\{V_\alpha\}$, $\{W_\alpha\}$ such that $\overline{V}_\alpha$ is compact and $\overline{W}_\alpha \subset V_\alpha$ for all α. [There is a sequence $\{S_n\}$ of precompact open sets such that $\overline{S}_n \subset S_{n+1}$ for all n and $X = \bigcup_1^\infty S_n$. For each n, $\{U \cap (S_{n+2} \setminus \overline{S}_{n-1}): U \in \mathscr{U}\}$ is an open cover of $\overline{S}_{n+1} \setminus S_n$. Choose a finite subcover to obtain $\{V_\alpha\}$ and mimic the beginning of the proof of Proposition (4.41) to obtain $\{W_\alpha\}$.]

(b) If X is a σ-compact LCH space, for any open cover $\mathscr{U}$ of X there is a partition of unity on X subordinate to $\mathscr{U}$ consisting of compactly supported functions.

6. TWO COMPACTNESS THEOREMS

The geometric objects on which one does analysis (Euclidean spaces, manifolds, etc.) tend to be compact or locally compact. However, in infinite-dimensional spaces such as spaces of functions, compactness is a rather rare

phenomenon and is much to be prized when it is available. Almost all compactness results in such situations are obtained via two basic theorems, Tychonoff's theorem and the Ascoli–Arzelà theorem, which we present in this section. Our proof of Tychonoff's theorem is based on the following result, which is of independent interest.

(4.42) Alexander's Lemma. Let X be a topological space whose topology is generated by a family $\mathscr{E}$ of sets. If every cover of X by members of $\mathscr{E}$ has a finite subcover, then X is compact.

Proof. Suppose X is not compact, and let $\mathfrak{A}$ be the collection of all open covers of X with no finite subcovers. $\mathfrak{A}$ is partially ordered by inclusion. If $\{\mathscr{A}_\beta\}_{\beta\in B}$ is a linearly ordered subcollection of $\mathfrak{A}$ and $U_1,\dots,U_n \in \bigcup_{\beta\in B}\mathscr{A}_\beta$, then $U_1,\dots,U_n \in \mathscr{A}_{\beta_0}$ for some $\beta_0 \in B$, so $\bigcup_1^n U_j \neq X$. It follows that $\bigcup_{\beta\in B}\mathscr{A}_\beta \in \mathfrak{A}$, so by Zorn's lemma $\mathfrak{A}$ has a maximal element $\mathscr{A}$. $\mathscr{A}$ is thus an open cover of X with no finite subcover, and if U is open and $U \notin \mathscr{A}$, $\mathscr{A}\cup\{U\}$ has a finite subcover. Let $\mathscr{B} = \mathscr{A}\cap\mathscr{E}$. We claim that $\mathscr{B}$ covers X; since no finite subfamily of $\mathscr{B}$ covers X, this contradicts our hypothesis.

Suppose to the contrary that there is a point $x \in X\setminus\bigcup\mathscr{B}$. (Here, $\bigcup\mathscr{B}$ is short for $\bigcup_{B\in\mathscr{B}}B$.) Choose $U\in\mathscr{A}$ with $x\in U$. Since $\mathscr{E}$ generates the topology, there exist $V_1,\dots,V_n\in\mathscr{E}$ with $x\in\bigcap_1^n V_j\subset U$. None of the V_j's is in $\mathscr{A}$ since $x\notin\bigcup\mathscr{B}$, so by maximality of $\mathscr{A}$, for each j there is a set W_j which is a finite union of sets in $\mathscr{A}$ such that $V_j\cup W_j = X$. But then

$$U\cup\left(\bigcup_1^n W_j\right)\supset\left(\bigcap_1^n V_j\right)\cup\left(\bigcup_1^n W_j\right) = X,$$

so that $\mathscr{A}$ has a finite subcover, contrary to assumption. □

(4.43) Tychonoff's Theorem. If $\{X_\alpha\}_{\alpha\in A}$ is any family of compact topological spaces, then $X = \prod_{\alpha\in A}X_\alpha$ (with the product topology) is compact.

Proof. The product topology is generated by the sets $\pi_\alpha^{-1}(U)$ where π_α: $X\to X_\alpha$ is the projection, U is open in X_α, and $\alpha\in A$. By Alexander's lemma it will suffice to prove that every cover $\mathscr{V}$ of X by such sets has a finite subcover. For each $\alpha\in A$ let $\mathscr{V}_\alpha$ be the collection of all open $U\subset X_\alpha$ such that $\pi_\alpha^{-1}(U)\in\mathscr{V}$. There must exist some $\beta\in A$ such that $\mathscr{V}_\beta$ covers X_β; otherwise there would be a point $x\in X$ such that $\pi_\alpha(x)\notin\bigcup\mathscr{V}_\alpha$ for every α, which would mean that $x\notin\bigcup\mathscr{V}$. Since X_β is compact, there exist $U_1,\dots,U_n\in\mathscr{V}_\beta$ such that $\bigcup_1^n U_j = X_\beta$. But then $\pi_\beta^{-1}(U_j)\in\mathscr{V}$ and $\bigcup_1^n\pi_\beta^{-1}(U_j) = \pi_\beta^{-1}(X_\beta) = X$, so we are done. □

We now turn to the Ascoli–Arzelà theorem, which has to do with compactness in spaces of continuous mappings. There are several variants of this result, of which the theorem below contains two of the most useful; see also Exercise 61.

If X is a topological space and $\mathscr{F}\subset C(X)$, $\mathscr{F}$ is called **equicontinuous at** $x\in X$ if for every $\varepsilon>0$ there is a neighborhood U of x such that $|f(y)-f(x)|<\varepsilon$ for all $y\in U$ and all $f\in\mathscr{F}$, and $\mathscr{F}$ is called **equicontinuous** if it is equicontinuous at each $x\in X$. Also, $\mathscr{F}$ is said to be **pointwise bounded** if $\{f(x)\colon f\in\mathscr{F}\}$ is a bounded subset of $\mathbf{C}$ for each $x\in X$.

(4.44) The Ascoli–Arzelà Theorem. (a) Let X be a compact Hausdorff space. If $\mathscr{F}$ is an equicontinuous, pointwise bounded subset of $C(X)$, then $\mathscr{F}$ is totally bounded (in the uniform metric), and the closure of $\mathscr{F}$ in $C(X)$ is compact.

(b) Let X be a σ-compact LCH space. If $\{f_n\}$ is an equicontinuous, pointwise bounded sequence in $C(X)$, there exist a subsequence $\{f_{n_j}\}$ of $\{f_n\}$ and an $f\in C(X)$ such that $f_{n_j}\to f$ uniformly on compact sets.

Proof. (a) Suppose $\varepsilon>0$. Since $\mathscr{F}$ is equicontinuous, for each $x\in X$ there is an open neighborhood U_x of x such that $|f(y)-f(x)|<\varepsilon/4$ for all $y\in U_x$ and all $f\in\mathscr{F}$. Since X is compact, we can choose $x_1,\dots,x_n\in X$ such that $\bigcup_1^n U_{x_j} = X$. Then $\{f(x_j)\colon f\in\mathscr{F}, j=1,\dots,n\}$ is a bounded subset of $\mathbf{C}$, so there is a finite set $\{z_1,\dots,z_m\}\subset\mathbf{C}$ which is $(\varepsilon/4)$-dense in it—that is, each $f(x_j)$ $(f\in\mathscr{F}, j=1,\dots,n)$ is at a distance less than $\varepsilon/4$ from some z_k. Let $A=\{x_1,\dots,x_n\}$, $B=\{z_1,\dots,z_m\}$: then the set B^A of functions from A to B is finite. For each $\varphi\in B^A$ let

$$\mathscr{F}_\varphi = \left\{f\in\mathscr{F}\colon |f(x_j)-\varphi(x_j)|<\varepsilon/4 \text{ for } j=1,\dots,n\right\}.$$

Then clearly $\bigcup\{\mathscr{F}_\varphi\colon \varphi\in B^A\} = \mathscr{F}$, and we claim that each $\mathscr{F}_\varphi$ has diameter at most ε, so we obtain a finite ε-dense subset of $\mathscr{F}$ by picking one f from each nonempty $\mathscr{F}_\varphi$. To prove the claim, suppose $f,g\in\mathscr{F}_\varphi$. Since $|f-\varphi|<\varepsilon/4$ and $|g-\varphi|<\varepsilon/4$ on A we have $|f-g|<\varepsilon/2$ on A. If $x\in X$, we have $x\in U_{x_j}$ for some j, and then

$$|f(x)-g(x)|\le|f(x)-f(x_j)|+|f(x_j)-g(x_j)|+|g(x_j)-g(x)|<\varepsilon.$$

This shows that $\mathscr{F}$ is totally bounded. Since the closure of a totally bounded set is totally bounded and $C(X)$ is complete, part (a) is proved.

(b) By Proposition (4.39) there is a sequence $\{U_k\}$ of precompact open sets such that $X=\bigcup_1^\infty U_k$. By part (a) there is a subsequence $\{f_j^1 = f_{n_j}\}$ of $\{f_n\}$ which is uniformly Cauchy on $\overline{U}_1$. Proceeding inductively, for $k\in\mathbf{N}$ we obtain a subsequence $\{f_j^k\}$ of $\{f_j^{k-1}\}$ which is uniformly Cauchy on $\overline{U}_k$. Let $g_k = f_k^k$: then $\{g_k\}$ is a subsequence of $\{f_n\}$ which is (except for the first $k-1$ terms) a subsequence of $\{f_j^k\}_j$ and hence is uniformly Cauchy on each $\overline{U}_k$. Let $f=\lim g_k$. Then $f\in C(X)$ and $g_k\to f$ uniformly on compact sets by Propositions (4.38) and (4.40). □

EXERCISES

57. If $\mathscr{E}$ is a base for the topology on X and every cover of X by members of $\mathscr{E}$ has a finite subcover, then X is compact. (This follows from Alexander's lemma, but unlike the latter, it can be proved from scratch in about two lines.)

58. If $\{X_\alpha\}_{\alpha \in A}$ is a family of topological spaces of which infinitely many are noncompact, then every compact subset of $\prod_{\alpha \in A} X_\alpha$ is nowhere dense.

59. The product of finitely many locally compact spaces is locally compact.

60. The product of countably many sequentially compact spaces is sequentially compact. [Use the "diagonal trick" as in the proof of Theorem (4.44b).]

61. Formulate a version of the Ascoli–Arzelà theorem in which functions on X are replaced by mappings from X to a metric space Y. (Use the same proof, but modify the hypotheses.)

62. Rephrase part (b) of the Ascoli–Arzelà theorem in a form similar to that of part (a) by using the metric in Exercise 55c.

63. Let $K \in C([0,1] \times [0,1])$. For $f \in C([0,1])$ define $Tf(x) = \int_0^1 K(x,y)f(y)\,dy$. Then $Tf \in C([0,1])$, and $\{Tf\colon \|f\|_u \le 1\}$ is precompact in $C([0,1])$.

64. Let (X,ρ) be a metric space. A function $f \in C(X)$ is called **Hölder continuous of exponent α** ($\alpha > 0$) if the quantity

$$N_\alpha(f) = \sup\left\{\frac{|f(x)-f(y)|}{\rho(x,y)^\alpha}\colon x \ne y\right\}$$

is finite. If X is compact, $\{f \in C(X)\colon \|f\|_u \le 1 \text{ and } N_\alpha(f) \le 1\}$ is compact in $C(X)$.

65. Let U be an open subset of $\mathbf{C}$ and $\{f_n\}$ a sequence of holomorphic functions on U. If $\{f_n\}$ is uniformly bounded on compact subsets of U, there is a subsequence which converges uniformly to a holomorphic function on compact subsets of U. (Use the Cauchy integral formula to obtain equicontinuity.)

7. THE STONE–WEIERSTRASS THEOREM

In this section we prove a far-reaching generalization of the well-known theorem of Weierstrass to the effect that any continuous function on a compact interval $[a,b]$ is the uniform limit of polynomials on $[a,b]$. Throughout this section X will denote a compact Hausdorff space, and we equip the space $C(X)$ with the uniform metric.

A subset $\mathscr{A}$ of $C(X,\mathbf{R})$ or $C(X)$ is said to **separate points** if for every $x, y \in X$ with $x \ne y$ there exists $f \in \mathscr{A}$ such that $f(x) \ne f(y)$. $\mathscr{A}$ is called an **algebra** if it is a real (resp. complex) vector subspace of $C(X,\mathbf{R})$ [resp. $C(X)$] such that $fg \in \mathscr{A}$ whenever $f \in \mathscr{A}$ and $g \in \mathscr{A}$. If $\mathscr{A} \subset C(X,\mathbf{R})$, $\mathscr{A}$ is called a **lattice** if $\max(f,g)$ and $\min(f,g)$ are in $\mathscr{A}$ for all $f, g \in \mathscr{A}$. Since the algebra and lattice operations are continuous, one easily sees that if $\mathscr{A}$ is an algebra or a lattice, so is its closure $\bar{\mathscr{A}}$ in the uniform metric.

(4.45) The Stone–Weierstrass Theorem. Let X be a compact Hausdorff space. If $\mathscr{A}$ is a closed subalgebra of $C(X,\mathbf{R})$ which separates points, then either $\mathscr{A} = C(X,\mathbf{R})$ or $\mathscr{A} = \{f \in C(X,\mathbf{R})\colon f(x_0) = 0\}$ for some $x_0 \in X$. The first alternative holds iff $\mathscr{A}$ contains the constant functions.

The proof will require several lemmas. The first one, in effect, proves the theorem when X consists of two points, and the second one is a special case of the theorem for $X = [-1,1]$. After these two we return to the general case.

(4.46) Lemma. Consider $\mathbf{R}^2$ as an algebra under coordinatewise addition and multiplication. The only subalgebras of $\mathbf{R}^2$ are $\mathbf{R}^2$, $\{(0,0)\}$, and the linear spans of $(1,0)$, $(0,1)$, and $(1,1)$.

Proof. The subspaces of $\mathbf{R}^2$ listed above are evidently subalgebras. If $\mathscr{A} \subset \mathbf{R}^2$ is a nonzero algebra and $(0,0) \ne (a,b) \in \mathscr{A}$, then $(a^2,b^2) \in \mathscr{A}$. If $a \ne 0$, $b \ne 0$, and $a \ne b$, then (a,b) and (a^2,b^2) are linearly independent, so $\mathscr{A} = \mathbf{R}^2$. The other possibilities—$a \ne 0 = b$, $a = 0 \ne b$, and $a = b \ne 0$ for all nonzero $(a,b) \in \mathscr{A}$—give the other three subalgebras. □

(4.47) Lemma. For any $\varepsilon > 0$ there is a polynomial P on $\mathbf{R}$ such that $P(0) = 0$ and $||x| - P(x)| < \varepsilon$ for $x \in [-1,1]$.

Proof. Consider the Maclaurin series for $(1-t)^{1/2}$:

$$(1-t)^{1/2} = 1 + \sum_1^\infty \left(\frac{-1}{2}\right)\left(\frac{1}{2}\right)\cdots\left(\frac{2n-3}{2}\right)\frac{t^n}{n!}$$

$$= 1 - \sum_1^\infty c_n t^n$$

where $c_n > 0$. By the ratio test, this series converges for $|t| < 1$; a proof that its sum is actually $(1-t)^{1/2}$ is outlined in Exercise 66. Moreover, by the monotone convergence theorem (applied to counting measure on $\mathbf{N}$),

$$\sum_1^\infty c_n = \lim_{t\to 1} \sum_1^\infty c_n t^n = 1 - \lim_{t\to 1}(1-t)^{1/2} = 1,$$

from which it follows that the series $1 - \sum_1^\infty c_n t_k$ converges absolutely and uniformly on $[-1,1]$ to $(1-t)^{1/2}$. Therefore, given $\varepsilon > 0$, by taking a suitable partial sum of this series we obtain a polynomial Q such that $|(1-t)^{1/2} - Q(t)| < \varepsilon/2$ for $t \in [-1,1]$. Setting $t = 1 - x^2$ and $R(x) = Q(1-x^2)$, we obtain a polynomial R such that $\big||x| - R(x)\big| < \varepsilon/2$ for $x \in [-1,1]$. In particular, $|R(0)| < \varepsilon/2$, so if we set $P(x) = R(x) - R(0)$, P is a polynomial such that $P(0) = 0$ and $\big||x| - P(x)\big| < \varepsilon$ for $x \in [-1,1]$. □

(4.48) Lemma. If $\mathscr{A}$ is a closed subalgebra of $C(X, \mathbf{R})$, then $|f| \in \mathscr{A}$ whenever $f \in \mathscr{A}$, and $\mathscr{A}$ is a lattice.

Proof. If $f \in \mathscr{A}$ and $f \neq 0$, let $h = f/\|f\|_u$. Then h maps X into $[-1,1]$, so if P is as in Lemma (4.47) we have $\|\,|h| - P \circ h\|_u < \varepsilon$. Since $P(0) = 0$, P has no constant term, so $P \circ h \in \mathscr{A}$ because $\mathscr{A}$ is an algebra. Since $\mathscr{A}$ is closed and ε is arbitrary, we have $|h| \in \mathscr{A}$ and hence $|f| = \|f\|_u |h| \in \mathscr{A}$. This proves the first assertion, and the second one follows because

$$\max(f,g) = \tfrac{1}{2}(f + g + |f - g|), \qquad \min(f,g) = \tfrac{1}{2}(f + g - |f - g|).$$

□

(4.49) Lemma. Suppose $\mathscr{A}$ is a closed lattice in $C(X, \mathbf{R})$. If $f \in C(X, \mathbf{R})$ and for every $x, y \in X$ there exists $g_{xy} \in \mathscr{A}$ such that $f(x) = g_{xy}(x)$ and $f(y) = g_{xy}(y)$, then $f \in \mathscr{A}$.

Proof. Given $\varepsilon > 0$, for each $x, y \in X$ let $U_{xy} = \{z \in X\colon f(z) < g_{xy}(z) + \varepsilon\}$ and $V_{xy} = \{z \in X\colon f(z) > g_{xy}(z) - \varepsilon\}$. These sets are open and contain x and y. Fix y: then $\{U_{xy}\}_{x \in X}$ covers X, so there is a finite subcover $\{U_{x_j y}\}_1^n$. Let $g_y = \max(g_{x_1 y}, \ldots, g_{x_n y})$: then $f < g_y + \varepsilon$ on X, and $f > g_y - \varepsilon$ on $V_y = \bigcap_1^n V_{x_j y}$, which is open and contains y. Thus $\{V_y\}_{y \in X}$ is another open cover of X, so there is a finite subcover $\{V_{y_j}\}_1^m$. Let $g = \min(g_{y_1}, \ldots, g_{y_m})$: then $\|f - g\|_u < \varepsilon$. Since $\mathscr{A}$ is a lattice, $g \in \mathscr{A}$; since $\mathscr{A}$ is closed and ε is arbitrary, $f \in \mathscr{A}$. □

Proof of Theorem (4.45). Given $x \neq y \in X$, let $\mathscr{A}_{xy} = \{(f(x), f(y))\colon f \in \mathscr{A}\}$. Then $\mathscr{A}_{xy}$ is a subalgebra of $\mathbf{R}^2$ in the sense of Lemma (4.46) because $f \to (f(x), f(y))$ is an algebra homomorphism. If $\mathscr{A}_{xy} = \mathbf{R}^2$ for all x, y, then Lemmas (4.48) and (4.49) imply that $\mathscr{A} = C(X, \mathbf{R})$. Otherwise, there exist x, y for which $\mathscr{A}_{xy}$ is a proper subalgebra of $\mathbf{R}^2$. It cannot be $\{(0,0)\}$ or the linear span of $(1,1)$ because $\mathscr{A}$ separates points, so by Lemma (4.46) $\mathscr{A}_{xy}$ is the linear span of $(1,0)$ or $(0,1)$. In either case there exists $x_0 \in X$ such that $f(x_0) = 0$ for all $f \in \mathscr{A}$. There is only one such x_0 since $\mathscr{A}$ separates points, so if neither x nor y is x_0, we have $\mathscr{A}_{xy} = \mathbf{R}^2$. Lemmas (4.48) and (4.49) now imply that $\mathscr{A} = \{f \in C(X, \mathbf{R})\colon f(x_0) = 0\}$. If $\mathscr{A}$ contains constant functions this cannot be the case, so $\mathscr{A}$ then equals $C(X, \mathbf{R})$. □

(4.50) Corollary (The Weierstrass Approximation Theorem). If X is a compact subset of $\mathbf{R}^n$, the restrictions to X of the (real) polynomial functions on $\mathbf{R}^n$ are dense in $C(X, \mathbf{R})$.

Proof. The set of polynomials is an algebra which separates points and contains constants. Its closure is therefore $C(X, \mathbf{R})$. □

The Stone–Weierstrass theorem, as it stands, is false for complex-valued functions. For example, the algebra of polynomials in one complex variable is not dense in $C(K)$ for most compact subsets K of $\mathbf{C}$. (In particular, if $K^o \neq \varnothing$, any uniform limit of polynomials on K must be holomorphic on K^o.) Here we shall give a simple proof that the function $f(z) = \bar{z}$ cannot be approximated uniformly by polynomials on the unit circle $\{e^{it}\colon 0 \le t \le 2\pi\}$. If $P(z) = \sum_0^n a_j z^j$, then

$$\int_0^{2\pi} \bar{f}(e^{it}) P(e^{it})\, dt = \sum_0^n a_j \int_0^{2\pi} e^{i(j+1)t}\, dt = 0.$$

Thus, abbreviating $f(e^{it})$ by f, etc., since $|f| = 1$ on the unit circle we have

$$2\pi = \left|\int_0^{2\pi} f\bar{f}\, dt\right| \le \left|\int_0^{2\pi} (f - P)\bar{f}\, dt\right| + \left|\int_0^{2\pi} \bar{f} P\, dt\right|$$

$$= \left|\int_0^{2\pi} (f - P)\bar{f}\, dt\right| \le \int_0^{2\pi} |f - P|\, dt \le 2\pi \|f - P\|_u.$$

Therefore, $\|f - P\|_u \ge 1$ for any polynomial $P(z)$.

There is, however, a complex version of the Stone–Weierstrass theorem:

(4.51) The Complex Stone–Weierstrass Theorem. Let X be a compact Hausdorff space. If $\mathscr{A}$ is a complex subalgebra of $C(X)$ which separates points and is closed under complex conjugation, then either $\mathscr{A} = C(X)$ or $\mathscr{A} = \{f \in C(X)\colon f(x_0) = 0\}$ for some $x_0 \in X$.

Proof. Since $\operatorname{Re} f = (f + \bar{f})/2$ and $\operatorname{Im} f = (f - \bar{f})/2i$, the set $\mathscr{A}_{\mathbf{R}}$ of real and imaginary parts of functions in $\mathscr{A}$ is a subalgebra of $C(X, \mathbf{R})$ to which the Stone–Weierstrass theorem applies. Since $\mathscr{A} = \{f + ig\colon f, g \in \mathscr{A}_{\mathbf{R}}\}$, the desired result follows. □

There is also a version of the Stone–Weierstrass theorem for noncompact LCH spaces. We state this result for real functions; the corresponding result for complex functions follows as in Theorem (4.51).

(4.52) Theorem. Let X be a noncompact LCH space. If $\mathscr{A}$ is a closed subalgebra of $C_0(X, \mathbf{R})$ $[= C_0(X) \cap C(X, \mathbf{R})]$ which separates points, then either $\mathscr{A} = C_0(X, \mathbf{R})$ or $\mathscr{A} = \{f \in C_0(X, \mathbf{R})\colon f(x_0) = 0\}$ for some $x_0 \in X$.

The proof is outlined in Exercise 67.

EXERCISES

66. Let $1 - \sum_1^\infty c_n t^n$ be the Maclaurin series for $(1-t)^{1/2}$.

(a) The series converges absolutely and uniformly on compact subsets of $(-1,1)$, as does the termwise differentiated series $-\sum_1^\infty n c_n t^{n-1}$. Thus if $f(t) = 1 - \sum_1^\infty c_n t^n$, then $f'(t) = -\sum_1^\infty n c_n t^{n-1}$, $-1 < t < 1$.

(b) By explicit calculation, $f(t) = -2(1-t)f'(t)$, from which it follows that $(1-t)^{-1/2} f(t)$ is constant. Also, $f(0) = 1$, so $f(t) = (1-t)^{1/2}$.

67. Prove Theorem (4.52). [If there exists $x_0 \in X$ such that $f(x_0) = 0$ for all $f \in \mathscr{A}$, let Y be the one-point compactification of $X \setminus \{x_0\}$; otherwise let Y be the one-point compactification of X. Apply Proposition (4.36) and the Stone–Weierstrass theorem on Y.]

68. Let X and Y be compact Hausdorff spaces. The algebra generated by functions of the form $f(x,y) = g(x)h(y)$, where $g \in C(X)$ and $h \in C(Y)$, is dense in $C(X \times Y)$.

69. Let A be a nonempty set, and let $X = [0,1]^A$. The algebra generated by the projections π_α: $X \to [0,1]$ ($\alpha \in A$) and the constant function 1 is dense in $C(X)$.

70. Let X be a compact Hausdorff space. An **ideal** in $C(X, \mathbf{R})$ is a subalgebra $\mathscr{I}$ of $C(X, \mathbf{R})$ such that if $f \in \mathscr{I}$ and $g \in C(X, \mathbf{R})$, then $fg \in \mathscr{I}$.

(a) If $\mathscr{I}$ is an ideal in $C(X, \mathbf{R})$, let $h(\mathscr{I}) = \{x \in X\colon f(x) = 0 \text{ for all } f \in \mathscr{I}\}$. Then $h(\mathscr{I})$ is a closed subset of X, called the **hull** of $\mathscr{I}$.

(b) If $E \subset X$, let $k(E) = \{f \in C(X, \mathbf{R})\colon f(x) = 0 \text{ for all } x \in E\}$. Then $k(E)$ is a closed ideal in $C(X, \mathbf{R})$, called the **kernel** of E.

(c) If $E \subset X$, then $h(k(E)) = \bar{E}$.

(d) If $\mathscr{I}$ is an ideal in $C(X, \mathbf{R})$, then $k(h(\mathscr{I})) = \bar{\mathscr{I}}$. [Hint: $\mathscr{I}$ may be identified with a subalgebra of $C_0(U, \mathbf{R})$ where $U = (h(\mathscr{I}))^c$.]

(e) The closed subsets of X are in one-to-one correspondence with the closed ideals of $C(X, \mathbf{R})$.

71. (This is a variation on the theme of Exercise 70; it does not use the Stone–Weierstrass theorem.) Let X be a compact Hausdorff space, and let M be the set of all nonzero algebra homomorphisms from $C(X, \mathbf{R})$ to $\mathbf{R}$. If $x \in X$, define $\hat{x} \in M$ by $\hat{x}(f) = f(x)$.

(a) If $\varphi \in M$, then $\{f \in C(X, \mathbf{R})\colon \varphi(f) = 0\}$ is a maximal proper ideal in $C(X, \mathbf{R})$.

(b) If $\mathscr{I}$ is a proper ideal in $C(X, \mathbf{R})$, there exists $x_0 \in X$ such that $f(x_0) = 0$ for all $f \in \mathscr{I}$. (Suppose not: construct an $f \in \mathscr{I}$ with $f > 0$ everywhere and conclude that $1 \in \mathscr{I}$. This requires no hard theorems.)

(c) The map $x \to \hat{x}$ is a bijection from X to M.

(d) If M is equipped with the topology of pointwise convergence, the map $x \to \hat{x}$ is a homeomorphism from X to M.

Conclusion: the topological structure of X is completely determined by the algebraic structure of $C(X, \mathbf{R})$, since M is defined purely algebraically.

8. EMBEDDINGS IN CUBES

We now present a technique for embedding topological spaces in products of intervals and discuss some of its applications. Throughout this section we shall denote the unit interval $[0,1]$ by I, and if A is any nonempty set, we shall call the product space I^A a **cube**.

If X is a topological space and $\mathscr{F} \subset C(X, I)$, we say that $\mathscr{F}$ **separates points and closed sets** if for every closed $E \subset X$ and every $x \in E^c$ there exists $f \in \mathscr{F}$ such that $f(x) \notin \overline{f(E)}$. If $\mathscr{F}$ separates points and closed sets, there is another family $\mathscr{G} \subset C(X, I)$ with the following slightly stronger property: for every closed set $E \subset X$ and every $x \in E^c$ there exists $g \in \mathscr{G}$ with $g(x) = 1$ and $g = 0$ on E. [If $f \in \mathscr{F}$ satisfies $f(x) \notin \overline{f(E)}$, take $g = h \circ f$ where $h \in C(I, I)$, $h(f(x)) = 1$, and $h = 0$ on $\overline{f(E)}$.] It follows that X admits a family $\mathscr{F}$ which separates points and closed sets iff X is completely regular.

Each $\mathscr{F} \subset C(X, I)$ canonically induces a map e: $X \to I^{\mathscr{F}}$ by the formula $\pi_f(e(x)) = f(x)$, where π_f: $I^{\mathscr{F}} \to I$ is the projection. We call e the map from X into the cube $I^{\mathscr{F}}$ **associated to** $\mathscr{F}$. (Evidently this construction can be generalized to target spaces other than I: cf. Exercise 20.)

(4.53) Proposition. Let X be a topological space, $\mathscr{F} \subset C(X, I)$, and e: $X \to I^{\mathscr{F}}$ the map associated to $\mathscr{F}$. Then:

(a) e is continuous.

(b) If $\mathscr{F}$ separates points, then e is injective.

(c) If $\mathscr{F}$ separates points and closed sets, then e is a homeomorphism onto its range.

Proof. (a) follows from Proposition (4.11), and (b) is obvious. To prove (c), suppose U is open in X. If $x \in U$, choose $f \in \mathscr{F}$ with $f(x) \notin \overline{f(U^c)}$ and let

$$V = \pi_f^{-1}\left[\overline{f(U^c)}\right]^c = \left\{p \in I^{\mathscr{F}}\colon \pi_f(p) \notin \overline{f(U^c)}\right\}.$$

Then V is open in $I^{\mathscr{F}}$ and

$$e(x) \in V \cap e(X) \subset e(U).$$

Thus $e(U)$ is a neighborhood of $e(x)$ in $e(X)$ for every $x \in U$, so $e(U)$ is open in $e(X)$ and e^{-1} is continuous. □

(4.54) Corollary. Every compact Hausdorff space is homeomorphic to a closed subset of a cube.

Proof. By Proposition (4.25) and Urysohn's lemma we can take $\mathscr{F} = C(X, I)$. □

(4.55) Corollary. A topological space is completely regular iff it is homeomorphic to a subset of a compact Hausdorff space.

Proof. Proposition (4.53) [with $\mathscr{F} = C(X, I)$] gives the "only if" implication; the converse is left to the reader (Exercise 72). □

Let us examine these results more closely. A **compactification** of a topological space X is a pair (Y, φ) where Y is a compact space and φ is a homeomorphism from X onto a dense subset of Y. (Frequently one identifies X with its image $\varphi(X) \subset Y$ and then speaks simply of "the compactification Y of X.") For example, $([-1, 1], \tanh)$ is a compactification of $\mathbf{R}$, and the one-point compactification (X^*, i) of an LCH space X is a compactification in the present sense, where i: $X \to X^*$ is the inclusion map. If X is completely regular and $\mathscr{F} \subset C(X, I)$ separates points and closed sets, let e: $X \to I^{\mathscr{F}}$ be the associated embedding and let Y be the closure of $e(X)$ in $I^{\mathscr{F}}$: then (Y, e) is a compactification of X. (Y, e) has the interesting property that if we identify X with its image $e(X)$, every $f \in \mathscr{F}$ has a continuous extension to Y (and this extension is unique, since X is dense in Y). Indeed, the identification of X with $e(X)$ simply turns f into the projection $\pi_f | e(X)$, which extends to $\pi_f | Y$. Moreover, if f and g are bounded continuous functions on X which extend continuously to Y, obviously so are $f + g$ and fg, and if $\{f_n\}$ is a uniformly convergent sequence of functions on X which extend continuously to Y, their extensions converge uniformly on Y since X is dense in Y, so $f = \lim f_n$ also extends continuously. We have proved:

(4.56) Proposition. Suppose $\mathscr{F} \subset C(X, I)$ separates points and closed sets. Let (Y, e) be the compactification of X associated to $\mathscr{F}$, and let $\mathscr{A}$ be the smallest uniformly closed subalgebra of $BC(X)$ containing $\mathscr{F}$. Then every $f \in \mathscr{A}$ has a unique continuous extension to Y.

This result has a converse: see Exercise 73.

If X is a completely regular space, the compactification of X associated to $\mathscr{F} = C(X, I)$ is called the **Stone–Čech compactification** of X and is denoted by $(\beta X, e)$, or simply by βX if we identify X with $e(X)$. Every $f \in BC(X)$ extends continuously to βX; in fact, a much more general result holds:

(4.57) Theorem. If X is a completely regular space, Y is a compact Hausdorff space, and $\varphi \in C(X, Y)$, then φ has a unique continuous extension $\tilde{\varphi}$ to βX—that is, there is a unique $\tilde{\varphi} \in C(\beta X, Y)$ such that $\tilde{\varphi} \circ e = \varphi$. If (Y, φ) is a compactification of X, then $\tilde{\varphi}$ is surjective; if also every $f \in BC(X)$ extends continuously to Y, then $\tilde{\varphi}$ is a homeomorphism.

Proof. Let $\mathscr{F} = C(X, I)$, $\mathscr{G} = C(Y, I)$, and let $(\beta Y, i)$ be the Stone-Čech compactification of Y. (Since Y is compact, i is a homeomorphism from Y to $\beta Y \subset I^{\mathscr{G}}$.) Given $\varphi \in C(X, Y)$, define Φ: $I^{\mathscr{G}} \to I^{\mathscr{F}}$ by $\pi_g(\Phi(p)) = \pi_{g \circ \varphi}(p)$. Φ is continuous by Proposition (4.11), and

$$\pi_g(\Phi(e(x))) = \pi_{g \circ \varphi}(e(x)) = \pi_g(i(\varphi(x))).$$

That is, $\Phi \circ e = i \circ \varphi$. It follows that $\Phi(e(X)) = i(\varphi(X)) \subset i(Y)$ and hence $\Phi(\beta X) \subset \overline{i(Y)} = \beta Y$. The situation is summarized in the following commutative diagram:

$$\begin{array}{ccccc} X & \xrightarrow{e} & \beta X & \subset & I^{\mathscr{F}} \\ \downarrow{\scriptstyle \varphi} & & \downarrow{\scriptstyle \Phi|\beta X} & & \downarrow{\scriptstyle \Phi} \\ Y & \xrightarrow{i} & \beta Y & \subset & I^{\mathscr{G}} \end{array}$$

Let $\tilde{\varphi} = i^{-1} \circ (\Phi | \beta X)$. Then $\tilde{\varphi} \circ e = i^{-1} \circ \Phi \circ e = \varphi$, and uniqueness of $\tilde{\varphi}$ is clear since $e(X)$ is dense in βX; thus the first assertion is proved. If (Y, φ) is a compactification of X, then $\varphi(X)$ is dense in Y; but then $\tilde{\varphi}(\beta X)$ is dense in Y and also compact, so that $\tilde{\varphi}(\beta X) = Y$. Finally, if every $f \in BC(X)$ extends continuously to Y, then every $f \in \mathscr{F}$ is of the form $g \circ \varphi$ for some $g \in \mathscr{G}$. In this case Φ is clearly injective; hence $\tilde{\varphi}$ is bijective and therefore [by Proposition (4.28)] a homeomorphism. □

This theorem shows that βX is the "largest" compactification of a completely regular space X, in the sense that every other compactification is a continuous image of it. At the other end of the scale, if X is locally compact, then $\mathscr{F} = C_c(X) \cap C(X, I)$ separates points and closed sets, by Theorem (4.32). A glance at the construction of the compactification (Y, e) associated to this $\mathscr{F}$ shows that Y consists of $e(X)$ together with the single point in $I^{\mathscr{F}}$ all of whose coordinates are zero. It is then easy to verify that Y is homeomorphic to the one-point compactification of X constructed in Section 5.

As a final application of the embedding e: $X \to I^{\mathscr{F}}$, we give a partial answer to the question: when is a topological space metrizable, that is, when is its topology defined by a metric? A necessary condition for X to be metrizable is that X be normal (Exercise 3). On the other hand,

(4.58) The Urysohn Metrization Theorem. Every second countable normal space is metrizable.

Since every subset of a metrizable space is metrizable (with the same metric), this result is an immediate consequence of Proposition (4.53) and the following two facts, whose proofs are outlined in Exercises 75 and 76:

(i) If X is normal and second countable, there is a countable family $\mathscr{F} \subset C(X, I)$ which separates points and closed sets.

(ii) If $\mathscr{F}$ is countable, $I^{\mathscr{F}}$ is metrizable.

EXERCISES

72. Every subset of a completely regular space is completely regular in the relative topology.

73. If X is a completely regular space, a subalgebra $\mathscr{A}$ of $BC(X)$ is called **completely regular** if (i) it is uniformly closed and contains constant functions, and (ii) $\mathscr{A} \cap C(X, I)$ separates points and closed sets.

(a) If (Y, e) is a Hausdorff compactification of X, $\mathscr{A}_Y = \{f \circ e\colon f \in C(Y)\}$ is a completely regular subalgebra of $BC(X)$.

(b) If (Y, e) and (Y', e') are Hausdorff compactifications of X such that $\mathscr{A}_Y = \mathscr{A}_{Y'}$, there is a homeomorphism $\varphi\colon Y \to Y'$ such that $\varphi \circ e = e'$. [Adapt the proof of Theorem (4.57), which deals with the case $Y = \beta X$.]

(c) If (Y, e) is the compactification of X associated to $\mathscr{F} \subset C(X, I)$, then $\mathscr{A}_Y$ is the smallest closed subalgebra of $BC(X)$ containing $\mathscr{F}$. (Use Exercise 69.)

(d) The Hausdorff compactifications of X are in one-to-one correspondence with the completely regular subalgebras of $BC(X)$.

74. Consider $\mathbf{N}$ (with the discrete topology) as a subset of its Stone-Čech compactification $\beta\mathbf{N}$.

(a) If A and B are disjoint subsets of $\mathbf{N}$, their closures in $\beta\mathbf{N}$ are disjoint. [Hint: $\chi_A \in C(\mathbf{N}, I)$.]

(b) No sequence in $\mathbf{N}$ converges in $\beta\mathbf{N}$ unless it is eventually constant (so $\beta\mathbf{N}$ is emphatically *not* sequentially compact).

75. If X is normal and second countable, there is a countable family $\mathscr{F} \subset C(X, I)$ which separates points and closed sets. [Let $\mathscr{B}$ be a countable base for the topology. Consider the set of pairs $(U, V) \in \mathscr{B} \times \mathscr{B}$ such that $\overline{U} \subset V$, and use Urysohn's lemma.]

76. Let $\{X_n\}_1^\infty$ be a sequence of metric spaces with metrics $\{\rho_n\}_1^\infty$ which take values in $[0, 1]$. (The last restriction can always be satisfied: see Exercise 55a.) If $x, y \in X = \prod_1^\infty X_n$, say $x = (x_1, x_2, \dots)$, $y = (y_1, y_2, \dots)$, define $\rho(x, y) = \sum_1^\infty 2^{-n}\rho_n(x_n, y_n)$. Then ρ is a metric which defines the product topology on X.

9. NOTES AND REFERENCES

The germ of the concept of topological space is clearly present in Riemann's lecture [79] on the foundations of geometry, delivered in 1854, but another half century passed before the mathematical world was ready to consider abstract spaces in a systematic way. The first attempt to construct an abstract framework for the study of limits and continuity was made in 1906 by Fréchet [35], who introduced metric spaces as well as a more general class of quasi-topological spaces whose properties were defined in terms of sequential convergence. A few years later Hausdorff [47] devised axioms for neighborhoods of points which amount to the definition of a Hausdorff space, and he deduced from them many of the basic results of general topology. The usefulness of his point of view was quickly recognized, and it became the foundation for the further development of the subject.

There are several good books to which the reader may refer for a more comprehensive treatment of point set topology, including Bourbaki [14], Dugundji [25], Engelking [28], Kelley [58], and Nagata [69]. Engelking [28] contains extensive references and historical notes.

Section 2. Examples of completely regular spaces which are not normal and regular spaces which are not completely regular may be found in the books listed above; they are all somewhat complicated. Particularly noteworthy is the existence of a regular space which admits no nonconstant continuous functions, a result due to Hewitt [51].

Section 3. The theory of nets is sometimes called the Moore–Smith theory of convergence, after its originators. Another general theory of convergence, invented by H. Cartan and publicized by Bourbaki, can be based on the notion of filters. A **filter** in a set X is a family $\mathscr{F} \subset \mathscr{P}(X)$ such that (i) if $F \in \mathscr{F}$ and $E \supset F$, then $E \in \mathscr{F}$; (ii) if $E, F \in \mathscr{F}$, then $E \cap F \in \mathscr{F}$; and (iii) $\varnothing \notin \mathscr{F}$. If X is a topological space, a filter $\mathscr{F}$ in X **converges** to $x \in X$ if every neighborhood of x belongs to $\mathscr{F}$. Filters and nets are related as follows. If $\langle x_\alpha \rangle_{\alpha \in A}$ is a net, its **derived filter** is the collection of all $E \subset X$ such that $\langle x_\alpha \rangle$ is eventually in E. On the other hand, if $\mathscr{F}$ is a filter, then $\mathscr{F}$ is a directed set under reverse inclusion, and a net $\langle x_F \rangle_{F \in \mathscr{F}}$ indexed by $\mathscr{F}$ is said to be **associated** to $\mathscr{F}$ if $x_F \in F$ for all $F \in \mathscr{F}$. It is then easy to verify that a net $\langle x_\alpha \rangle_{\alpha \in A}$ converges to x iff its derived filter converges to x, and a filter $\mathscr{F}$ converges to x iff all of its associated nets converge to x.

Section 4. The usage of the term "compact" is not completely standardized. In some older works the term "bicompact" is used instead, and some authors use "compact" to mean compact Hausdorff and "quasicompact" to mean compact. Synonyms for "precompact" which are frequently found in the literature are "conditionally compact" and "relatively compact"; the latter one

is infelicitous because it suggests compactness in the relative topology, which is quite different.

Section 6. The axiom of choice is used twice in proving Tychonoff's theorem: once in the form of Zorn's lemma in the proof of Alexander's lemma, and again in the assertion "otherwise there would be a point $x \in X\ldots$" in the final argument. It is an intriguing fact, first discovered by Kelley [57], that Tychonoff's theorem in turn implies the axiom of choice. Here is the proof: suppose $\{X_\alpha\}_{\alpha \in A}$ is a nonempty collection of nonempty sets. Pick a point ω which is not an element of any X_α, set $X_\alpha^* = X_\alpha \cup \{\omega\}$, and define a topology on X_α^* by declaring the open sets to be $\varnothing$, X_α, $\{\omega\}$, and X_α^*. Evidently X_α^* is compact, hence so is $X^* = \prod_{\alpha \in A} X_\alpha^*$, by Tychonoff's theorem. Let $F_\alpha = \pi_\alpha^{-1}(X_\alpha)$. The sets F_α are closed, and by the axiom of choice for finite collections of sets—which is provable from the other axioms of set theory—they have the finite intersection property. Indeed, given a finite set $B \subset A$, we can pick $x_\beta \in X_\beta$ for each $\beta \in B$; then $\bigcap_{\beta \in B} F_\beta$ contains the point $x \in X$ such that $\pi_\beta(x) = x_\beta$ for $\beta \in B$ and $\pi_\alpha(x) = \omega$ for $\alpha \notin B$. By Proposition (4.21), $\bigcap_{\alpha \in A} F_\alpha$, which is precisely $\prod_{\alpha \in A} X_\alpha$, is nonempty.

The original results of Ascoli and Arzelà had to do with functions on $\mathbf{R}$. Other versions of the Ascoli–Arzelà theorem, pertaining to the compactness of subsets of $C(X, Y)$ under various hypotheses on X and Y, can be found in the books listed above and in Royden [87].

Section 7. For a survey of the Stone–Weierstrass theorem and its ramifications, see Stone [103].

Section 8. It is not hard to show that every second countable regular space is normal (see Kelley [58], Lemma 4.1); consequently, the hypothesis of normality in the Urysohn metrization theorem can be replaced by regularity. Necessary and sufficient conditions are known for an arbitrary topological space to be metrizable, but they are not as readily verifiable as the conditions in Urysohn's theorem. See the references listed above.

Occasionally the term "compactification" is used to mean a continuous injection $\varphi\colon X \to Y$ from a topological space X onto a dense subset of a compact space Y without the requirement that it be a homeomorphism onto its range, particularly in the so-called "Bohr compactification" of a locally compact group.

5

ELEMENTS OF FUNCTIONAL ANALYSIS

"Functional analysis" is the traditional name for the study of infinite-dimensional vector spaces over $\mathbf{R}$ or $\mathbf{C}$ and the linear maps between them. What distinguishes this subject from mere linear algebra is the importance of topological considerations. On finite-dimensional vector spaces there is only one reasonable topology, and linear maps are automatically continuous, but in infinite dimensions things are not so simple. (As we have already observed, if $\{f_n\}$ is a sequence of functions on $\mathbf{R}$, there are many things one can mean by the statement "$f_n \to f$.") As our aim in this chapter is only to give a brief introduction to the subject, we shall restrict attention—except in Section 4—to topologies on vector spaces defined by norms.

1. NORMED VECTOR SPACES

Let K denote either $\mathbf{R}$ or $\mathbf{C}$, and let $\mathscr{X}$ be a vector space over K. By a **subspace** of $\mathscr{X}$ we shall always mean a vector subspace. If $x \in \mathscr{X}$, we denote by Kx the one-dimensional subspace spanned by x.

A **seminorm** on $\mathscr{X}$ is a function $x \to \|x\|$ from $\mathscr{X}$ to $[0, \infty)$ such that $\|x + y\| \le \|x\| + \|y\|$ for all $x, y \in \mathscr{X}$ (the **triangle inequality**) and $\|\lambda x\| = |\lambda|\|x\|$ for all $x \in \mathscr{X}$ and $\lambda \in K$. A seminorm such that $\|x\| = 0$ iff $x = 0$ is called a **norm,** and a vector space equipped with a norm is called a **normed vector space** (or **normed linear space**). If $\mathscr{X}$ is a normed vector space, the function $\rho(x, y) = \|x - y\|$ is clearly a metric on $\mathscr{X}$; the topology it defines is called the **norm topology** on $\mathscr{X}$. Two norms $\|\ \|_1$ and $\|\ \|_2$ on $\mathscr{X}$ are called **equivalent** if there exist $C_1, C_2 > 0$ such that

$$C_1\|x\|_1 \le \|x\|_2 \le C_2\|x\|_1 \quad (x \in \mathscr{X}).$$

Equivalent norms define the same topology and the same Cauchy sequences (Exercise 2).

A normed vector space which is complete with respect to the norm metric is called a **Banach space**. (Every normed vector space can be embedded in a Banach space as a dense subspace. One way to do this is to mimic the construction of $\mathbf{R}$ from $\mathbf{Q}$ via Cauchy sequences; we shall present a simpler way in Section 2.) The following is a useful criterion for completeness of a normed vector space $\mathscr{X}$. If $\{x_n\}$ is a sequence in $\mathscr{X}$, the series $\sum_1^\infty x_n$ is said to **converge** to x if $\sum_1^N x_n \to x$ as $N \to \infty$, and it is called **absolutely convergent** if $\sum_1^\infty \|x_n\| < \infty$.

(5.1) Theorem. A normed vector space $\mathscr{X}$ is complete iff every absolutely convergent series in $\mathscr{X}$ converges.

Proof. If $\mathscr{X}$ is complete and $\sum_1^\infty \|x_n\| < \infty$, let $S_N = \sum_1^N x_n$. Then for $N > M$,

$$\|S_N - S_M\| \le \sum_{M+1}^{\infty} \|x_n\| \to 0 \quad \text{as } M, N \to \infty,$$

so the sequence $\{S_N\}$ is Cauchy and hence convergent. Conversely, suppose every absolutely convergent series converges and $\{x_n\}$ is a Cauchy sequence. We can choose $n_1 < n_2 < \cdots$ such that $\|x_n - x_m\| < 2^{-j}$ for $m, n \ge n_j$. Let $y_1 = x_{n_1}$ and $y_j = x_{n_j} - x_{n_{j-1}}$ for $j > 1$. Then $\sum_1^k y_j = x_{n_k}$, and

$$\sum_1^\infty \|y_j\| \le \|y_1\| + \sum_1^\infty 2^{-j} = \|y_1\| + 1 < \infty,$$

so $\lim x_{n_k} = \sum_1^\infty y_j$ exists. But since $\{x_n\}$ is Cauchy, it is easily verified that $\{x_n\}$ converges to the same limit as $\{x_{n_k}\}$. □

We have already seen some examples of Banach spaces. First, if X is a topological space, $B(X)$ and $BC(X)$ are Banach spaces with the uniform norm. Second, if $(X, \mathscr{M}, \mu)$ is a measure space, $L^1(\mu)$ is a Banach space with the L^1 norm $\|f\|_1 = \int |f|\, d\mu$. [Observe that $\|\ \|_1$ is only a seminorm if we think of $L^1(\mu)$ as consisting of individual functions, but it becomes a norm if we identify two functions which are equal a.e.] That $L^1(\mu)$ is complete follows from Theorems (2.25) and (5.1): if $\sum_1^\infty \|f_n\|_1 < \infty$, Theorem (2.25) shows that $f = \sum_1^\infty f_n$ exists a.e., and

$$\int \left| f - \sum_1^N f_n \right| d\mu \le \sum_{N+1}^{\infty} \int |f_n|\, d\mu \to 0 \text{ as } N \to \infty.$$

More examples will be found in the exercises and in subsequent sections.

If $\mathscr{X}$ and $\mathscr{Y}$ are normed vector spaces, $\mathscr{X} \times \mathscr{Y}$ becomes a normed vector space when equipped with the **product norm** $\|(x, y)\| = \max(\|x\|, \|y\|)$. (Here, of course, $\|x\|$ refers to the norm on $\mathscr{X}$ while $\|y\|$ refers to the norm on $\mathscr{Y}$.) Sometimes other norms equivalent to this one, such as $\|(x, y)\| = \|x\| + \|y\|$ or $\|(x, y)\| = (\|x\|^2 + \|y\|^2)^{1/2}$ are used instead.

A linear map $T\colon \mathscr{X} \to \mathscr{Y}$ between two normed vector spaces is called **bounded** if there exists $C > 0$ such that $\|Tx\| \le C\|x\|$ for all $x \in \mathscr{X}$. This is different from the notion of boundedness for functions on a set, according to which T would be bounded if $\|Tx\| \le C$ for all x. Clearly no nonzero linear map can satisfy the latter condition, since $T(\lambda x) = \lambda Tx$ for all scalars λ. The present definition means that T is bounded on bounded subsets of $\mathscr{X}$.

(5.2) Proposition. If $\mathscr{X}$ and $\mathscr{Y}$ are normed vector spaces and $T\colon \mathscr{X} \to \mathscr{Y}$ is a linear map, the following are equivalent:

(a) T is continuous.

(b) T is continuous at 0.

(c) T is bounded.

Proof. That (a) implies (b) is trivial. If T is continuous at $0 \in \mathscr{X}$, there is a neighborhood U of 0 such that $T(U) \subset \{y \in \mathscr{Y}\colon \|y\| \le 1\}$, and U must contain a ball $B = \{x \in \mathscr{X}\colon \|x\| \le \delta\}$ about 0. Thus $\|Tx\| \le 1$ when $\|x\| \le \delta$, so for any $x \in \mathscr{X}$,

$$\|Tx\| = \left\| \frac{\|x\|}{\delta} T\left(\frac{\delta x}{\|x\|} \right) \right\| \le \frac{\|x\|}{\delta},$$

which shows that (b) implies (c). Finally, if $\|Tx\| \le C\|x\|$ for all x, then $\|Tx_1 - Tx_2\| = \|T(x_1 - x_2)\| \le \varepsilon$ whenever $\|x_1 - x_2\| \le \varepsilon/C$, so that T is continuous. □

If $\mathscr{X}$ and $\mathscr{Y}$ are normed vector spaces, we denote the space of all bounded linear maps from $\mathscr{X}$ to $\mathscr{Y}$ by $L(\mathscr{X}, \mathscr{Y})$. It is easily verified that $L(\mathscr{X}, \mathscr{Y})$ is a vector space and that the functional defined by

$$\|T\| = \sup\{\|Tx\|\colon \|x\| = 1\} = \sup\left\{ \frac{\|Tx\|}{\|x\|}\colon x \ne 0 \right\}$$

is a norm on $L(\mathscr{X}, \mathscr{Y})$, called the **operator norm**. We always assume $L(\mathscr{X}, \mathscr{Y})$ to be equipped with this norm unless we specify otherwise.

(5.3) Proposition. If $\mathscr{Y}$ is complete, so is $L(\mathscr{X}, \mathscr{Y})$.

Proof. Let $\{T_n\}$ be a Cauchy sequence in $L(\mathscr{X}, \mathscr{Y})$. If $x \in \mathscr{X}$, then $\{T_n x\}$ is Cauchy in $\mathscr{Y}$ because $\|T_n x - T_m x\| \le \|T_n - T_m\|\,\|x\|$. Define $T\colon \mathscr{X} \to \mathscr{Y}$ by

$Tx = \lim T_n x$. The reader may readily verify that $T \in L(\mathscr{X}, \mathscr{Y})$ (in fact, $\|T\| = \lim \|T_n\|$) and that $\|T_n - T\| \to 0$. □

Another useful property of the operator norm is the following. If $T \in L(\mathscr{X}, \mathscr{Y})$ and $S \in L(\mathscr{Y}, \mathscr{Z})$, then

$$\|STx\| \le \|S\| \|Tx\| \le \|S\| \|T\| \|x\|,$$

so that $ST \in L(\mathscr{X}, \mathscr{Z})$ and $\|ST\| \le \|S\| \|T\|$. In particular, $L(\mathscr{X}, \mathscr{X})$ is an algebra. If $\mathscr{X}$ is complete, $L(\mathscr{X}, \mathscr{X})$ is in fact a **Banach algebra:** a Banach space which is also an algebra, such that the norm of a product is at most the product of the norms. [Another example of a Banach algebra is $BC(X)$, where X is a topological space, under pointwise multiplication and the uniform norm.]

If $T \in L(\mathscr{X}, \mathscr{Y})$, T is said to be **invertible**, or an **isomorphism**, if T is bijective and T^{-1} is bounded (in other words, $\|Tx\| \ge C\|x\|$ for some $C > 0$). T is called an **isometry** if $\|Tx\| = \|x\|$ for all $x \in \mathscr{X}$. An isometry is injective but not necessarily surjective; it is, however, an isomorphism onto its range.

EXERCISES

1. If $\mathscr{X}$ is a normed vector space over K ($= \mathbf{R}$ or $\mathbf{C}$), then addition and scalar multiplication are continuous from $\mathscr{X} \times \mathscr{X}$ and $K \times \mathscr{X}$ to $\mathscr{X}$. Moreover, the norm is continuous from $\mathscr{X}$ to $[0, \infty)$; in fact, $|\|x\| - \|y\|| \le \|x - y\|$.

2. Equivalent norms define the same open sets and the same Cauchy sequences.

3. If $\mathscr{X}, \mathscr{Y}$ are normed vector spaces, the map $(T, x) \to Tx$ from $L(\mathscr{X}, \mathscr{Y}) \times \mathscr{X}$ to $\mathscr{Y}$ is continuous. (That is, if $T_n \to T$ and $x_n \to x$, then $T_n x_n \to Tx$.)

4. If $\mathscr{X}$ is finite-dimensional, all norms on $\mathscr{X}$ are equivalent. (Let $e_1, \dots, e_n$ be a basis for $\mathscr{X}$, and let $\|\sum_1^n a_j e_j\|_1 = \sum_1^n |a_j|$. Then $\| \|_1$ is a norm on $\mathscr{X}$; compare any other norm to it. It is useful to observe that $\{x \in \mathscr{X}: \|x\|_1 = 1\}$ is compact in the topology defined by $\| \|_1$.)

5. Let $\mathscr{X}$ be a Banach space.

(a) If $T \in L(\mathscr{X}, \mathscr{X})$ and $\|I - T\| < 1$ where I is the identity operator, then T is invertible. [The series $\sum_0^\infty (I - T)^n$ converges in $L(\mathscr{X}, \mathscr{X})$ to T^{-1}.]

(b) If $T \in L(\mathscr{X}, \mathscr{X})$ is invertible and $\|T - S\| \le \|T^{-1}\|^{-1}$, then S is invertible. Thus the set of invertible operators is open in $L(\mathscr{X}, \mathscr{X})$.

6. Let $(X, \mathscr{M})$ be a measurable space, and let $M(X)$ be the space of complex measures on $(X, \mathscr{M})$. Then $\|\mu\| = |\mu|(X)$ is a norm on $M(X)$ which makes $M(X)$ into a Banach space. [Use Theorem (5.1).]

7. Let $C^k([0,1])$ be the space of functions on $[0,1]$ possessing continuous derivatives up to order k on $[0,1]$ (including one-sided derivatives at the endpoints).

(a) If $f \in C([0,1])$, then $f \in C^k([0,1])$ iff f is k times continuously differentiable on $(0,1)$ and $f^{(j)}$ extends continuously to $[0,1]$ for $j \le k$. (The mean value theorem is useful.)

(b) $\|f\| = \sum_0^k \|f^{(j)}\|_u$ is a norm on $C^k([0,1])$ which makes $C^k([0,1])$ into a Banach space. [Use induction on k. The essential point is that if $\{f_n\} \subset C^1([0,1])$, $f_n \to f$ uniformly, and $f_n' \to g$ uniformly, then $f \in C^1([0,1])$ and $f' = g$. The easy way to prove this is to show that $f(x) - f(0) = \int_0^x g(t)\,dt$.]

8. Let $L_k^1([0,1])$ be the space of all $f \in C^{k-1}([0,1])$ such that $f^{(k-1)}$ is absolutely continuous on $[0,1]$ [and hence $f^{(k)}$ exists a.e. and is in $L^1([0,1])$]. Then $\|f\| = \sum_0^k \int_0^1 |f^{(j)}(x)|\,dx$ is a norm on $L_k^1([0,1])$ which makes $L_k^1([0,1])$ into a Banach space. (Cf. Exercise 7 and its hint.)

9. If $0 < \alpha \le 1$, let $\Lambda_\alpha([0,1])$ be the set of Hölder continuous functions of exponent α on $[0,1]$. Thus $f \in \Lambda_\alpha([0,1])$ iff $\|f\|_{(\alpha)} < \infty$, where

$$\|f\|_{(\alpha)} = |f(0)| + \sup_{x \ne y} \frac{|f(x) - f(y)|}{|x - y|^\alpha}.$$

(a) $\| \|_{(\alpha)}$ is a norm which makes $\Lambda_\alpha([0,1])$ into a Banach space.

(b) Let $\lambda_\alpha([0,1])$ be the set of all $f \in \Lambda_\alpha([0,1])$ such that

$$\frac{|f(x) - f(y)|}{|x - y|^\alpha} \to 0 \quad \text{as } x \to y, \quad \text{for all } y \in [0,1].$$

If $\alpha < 1$, $\lambda_\alpha([0,1])$ is a closed subspace of $\Lambda_\alpha([0,1])$. If $\alpha = 1$, $\lambda_\alpha([0,1])$ contains only constant functions.

The next five problems deal with quotient spaces. If $\mathscr{X}$ is a vector space and $\mathscr{M}$ is a subspace, define an equivalence relation on $\mathscr{X}$ by $x \sim y$ iff $x - y \in \mathscr{M}$, and denote the equivalence class of x by $x + \mathscr{M}$. The set of equivalence classes, denoted by $\mathscr{X}/\mathscr{M}$, is a vector space with vector operations $(x + \mathscr{M}) + (y + \mathscr{M}) = (x + y) + \mathscr{M}$, $\lambda(x + \mathscr{M}) = (\lambda x) + \mathscr{M}$.

10. If $\mathscr{X}$ is a normed vector space, the closure of any subspace is a subspace.

11. Let $\mathscr{X}$ be a normed vector space and $\mathscr{M}$ a closed subspace.

(a) $\|x + \mathscr{M}\| = \inf\{\|x + y\|: y \in \mathscr{M}\}$ is a norm on $\mathscr{X}/\mathscr{M}$, called the **quotient norm**.

(b) For any $\varepsilon > 0$ there exists $x \in \mathscr{X}$ such that $\|x\| = 1$ and $\|x + \mathscr{M}\| \ge 1 - \varepsilon$.

(c) The projection $\pi(x) = x + \mathscr{M}$ from $\mathscr{X}$ to $\mathscr{X}/\mathscr{M}$ has norm 1.

(d) If $\mathscr{X}$ is complete, so is $\mathscr{X}/\mathscr{M}$. [Use Theorem (5.1).]

12. If $\| \ \|$ is a seminorm on $\mathscr{X}$, let $\mathscr{M} = \{x \in X: \|x\| = 0\}$. Then $\mathscr{M}$ is a subspace, and $\|x + \mathscr{M}\| = \|x\|$ defines a norm on $\mathscr{X}/\mathscr{M}$.

13. If $\mathscr{X}$ is a normed vector space and $\mathscr{M}$ is a nonclosed subspace, then $\|x + \mathscr{M}\|$, as defined in Exercise 11, is a seminorm on $\mathscr{X}/\mathscr{M}$. If one divides by its nullspace as in Exercise 12, the resulting quotient space is isometrically isomorphic to $\mathscr{X}/\overline{\mathscr{M}}$.

14. Suppose $\mathscr{X}$ and $\mathscr{Y}$ are normed vector spaces and $T \in L(\mathscr{X}, \mathscr{Y})$.

(a) $\mathscr{N}(T) = \{x: Tx = 0\}$ is a closed subspace of $\mathscr{X}$.

(b) If $\mathscr{M}$ is a closed subspace of $\mathscr{N}(T)$, there is a unique $S \in L(\mathscr{X}/\mathscr{M}, \mathscr{Y})$ such that $T = S \circ \pi$ where $\pi: \mathscr{X} \to \mathscr{X}/\mathscr{M}$ is the projection. Moreover, $\|S\| = \|T\|$.

The next six exercises develop a theory of integration for functions with values in a separable Banach space. Let $(X, \mathscr{M}, \mu)$ be a measure space, $\mathscr{Y}$ a separable Banach space, $L_{\mathscr{Y}}$ the set of $(\mathscr{M}, \mathscr{B}_{\mathscr{Y}})$-measurable maps from X to $\mathscr{Y}$, and $F_{\mathscr{Y}}$ the set of maps $f: X \to \mathscr{Y}$ of the form $f = \sum_1^n y_j \chi_{E_j}$ where $y_j \in \mathscr{Y}$, $E_j \in \mathscr{M}$, and $\mu(E_j) < \infty$. If $f \in L_{\mathscr{Y}}$, since $y \to \|y\|$ is continuous (Exercise 1), $x \to \|f(x)\|$ is $(\mathscr{M}, \mathscr{B}_{\mathbf{R}})$-measurable, and we define $\|f\|_1 = \int \|f(x)\| \, d\mu(x)$. Finally, let $L^1_{\mathscr{Y}} = \{f \in L_{\mathscr{Y}}: \|f\|_1 < \infty\}$.

15. $L_{\mathscr{Y}}$ is a vector space, $F_{\mathscr{Y}}$ and $L^1_{\mathscr{Y}}$ are subspaces of it, $F_{\mathscr{Y}} \subset L^1_{\mathscr{Y}}$, and $\| \ \|_1$ is a seminorm on $L^1_{\mathscr{Y}}$ which becomes a norm if we identify two functions which are equal a.e.

16. Let $\{y_n\}_1^\infty$ be a countable dense set in $\mathscr{Y}$. Given $\varepsilon > 0$, let $B_n^\varepsilon = \{y \in \mathscr{Y}: \|y - y_n\| < \varepsilon \|y_n\|\}$. Then $\bigcup_1^\infty B_n^\varepsilon \supset (\mathscr{Y} \setminus \{0\})$.

17. If $f \in L^1_{\mathscr{Y}}$, there is a sequence $\{h_n\} \subset F_{\mathscr{Y}}$ with $h_n \to f$ a.e. and $\|h_n - f\|_1 \to 0$. [With notation as in Exercise 16, let $A_{nj} = B_n^{1/j} \setminus \bigcup_{m=1}^{n-1} B_m^{1/j}$ and $E_{nj} = f^{-1}(A_{nj})$, and consider $g_j = \sum_1^\infty y_n \chi_{E_{nj}}$.]

18. There is a unique linear map $\int: L^1_{\mathscr{Y}} \to \mathscr{Y}$ such that $\int y\chi_E = \mu(E)y$ for $y \in \mathscr{Y}$ and $E \in \mathscr{M}$ $(\mu(E) < \infty)$, and $\|\int f\| \le \|f\|_1$.

19. (The dominated convergence theorem) If $\{f_n\}$ is a sequence in $L^1_{\mathscr{Y}}$ and there exists $g \in L^1$ such that $\|f(x)\| \le g(x)$ for a.e. x, and $f_n \to f$ a.e., then $\int f_n \to \int f$.

20. If $\mathscr{Z}$ is a separable Banach space, $T \in L(\mathscr{Y}, \mathscr{Z})$, and $f \in L^1_{\mathscr{Y}}$, then $T \circ f \in L^1_{\mathscr{Z}}$ and $\int T \circ f = T \int f$.

2. LINEAR FUNCTIONALS

Let $\mathscr{X}$ be a vector space over K, where $K = \mathbf{R}$ or $\mathbf{C}$. A linear map from $\mathscr{X}$ to K is called a **linear functional** on $\mathscr{X}$. If $\mathscr{X}$ is a normed vector space, the space $L(\mathscr{X}, K)$ of bounded linear functionals on $\mathscr{X}$ is called the **dual space** of $\mathscr{X}$ and is denoted by $\mathscr{X}^*$. According to Proposition (5.3), $\mathscr{X}^*$ is a Banach space with the operator norm. If $\mathscr{X}$ is a vector space over $\mathbf{C}$, it is also a vector space over

$\mathbf{R}$, and we can consider both real and complex linear functionals on $\mathscr{X}$. The relationship between the two is as follows:

(5.4) Proposition. Let $\mathscr{X}$ be a vector space over $\mathbf{C}$. If f is a complex linear functional and $u = \operatorname{Re} f$, then u is a real linear functional, and $f(x) = u(x) - iu(ix)$ for all $x \in \mathscr{X}$. Conversely, if u is a real linear functional and f is defined by $f(x) = u(x) - iu(ix)$, then f is complex linear. In this case, if $\mathscr{X}$ is normed, we have $\|u\| = \|f\|$.

Proof. If f is complex linear and $u = \operatorname{Re} f$, u is clearly real linear and $\operatorname{Im} f(x) = -\operatorname{Re}(if(x)) = -u(ix)$, so $f(x) = u(x) - iu(ix)$. If u is real linear and $f(x) = u(x) - iu(ix)$, then f is clearly linear over $\mathbf{R}$, and $f(ix) = u(ix) - iu(-x) = u(ix) + iu(x) = if(x)$, so f is also linear over $\mathbf{C}$. Finally, if $\mathscr{X}$ is normed, since $|u(x)| = |\operatorname{Re} f(x)| \le |f(x)|$ we have $\|u\| \le \|f\|$. On the other hand, if $f(x) \neq 0$, let $\alpha = \operatorname{sgn} f(x)$. Then $|f(x)| = \alpha f(x) = f(\alpha x) = u(\alpha x)$ [since $f(\alpha x)$ is real], so $|f(x)| \le \|u\| \|\alpha x\| = \|u\| \|x\|$, whence $\|f\| \le \|u\|$. □

It is not obvious that there are any bounded linear functionals on an arbitrary normed vector space except the zero functional. The fact that such functionals exist in great abundance is one of the fundamental theorems of functional analysis. We shall now present this result in a somewhat more general form which has other important applications.

If $\mathscr{X}$ is a real vector space, a **Minkowski functional** on $\mathscr{X}$ is a map $p: \mathscr{X} \to \mathbf{R}$ such that $p(x + y) \le p(x) + p(y)$ and $p(\lambda x) = \lambda p(x)$ for all $x, y \in \mathscr{X}$ and all $\lambda \ge 0$. (For example, every seminorm is a Minkowski functional.)

(5.5) The Hahn–Banach Theorem. Let $\mathscr{X}$ be a real vector space, p a Minkowski functional on $\mathscr{X}$, $\mathscr{M}$ a subspace of $\mathscr{X}$, and f a linear functional on $\mathscr{M}$ such that $f(x) \le p(x)$ for all $x \in \mathscr{M}$. Then there exists a linear functional F on $\mathscr{X}$ such that $F(x) \le p(x)$ for all $x \in \mathscr{X}$ and $F|\mathscr{M} = f$.

Proof. We begin by showing that if $x \in \mathscr{X} \setminus \mathscr{M}$, f can be extended to a linear functional g on $\mathscr{M} + \mathbf{R}x$ satisfying $g(y) \le p(y)$ for $y \in \mathscr{M} + \mathbf{R}x$. If $y_1, y_2 \in \mathscr{M}$, we have

$$f(y_1) + f(y_2) = f(y_1 + y_2) \le p(y_1 + y_2) \le p(y_1 - x) + p(x + y_2),$$

or

$$f(y_1) - p(y_1 - x) \le p(x + y_2) - f(y_2).$$

Hence

$$\sup\{f(y) - p(y - x): y \in \mathscr{M}\} \le \inf\{p(x + y) - f(y): y \in \mathscr{M}\}.$$

Let α be any number satisfying

$$\sup\{f(y) - p(y - x): y \in \mathscr{M}\} \le \alpha \le \inf\{p(x + y) - f(y): y \in \mathscr{M}\}$$

and define g: $\mathscr{M} + \mathbf{R}x \to \mathbf{R}$ by $g(y + \lambda x) = f(y) + \lambda\alpha$. Then g is clearly linear, and $g|\mathscr{M} = f$, so that $g(y) \le p(y)$ for $y \in \mathscr{M}$. Moreover, if $\lambda > 0$ and $y \in \mathscr{M}$,

$$g(y + \lambda x) = \lambda\left[f\left(\frac{y}{\lambda}\right) + \alpha\right] \le \lambda\left[f\left(\frac{y}{\lambda}\right) + p\left(x + \left(\frac{y}{\lambda}\right)\right) - f\left(\frac{y}{\lambda}\right)\right]$$

$$= p(y + \lambda x),$$

whereas if $\lambda < 0$,

$$g(y + \lambda x) = |\lambda|\left[f\left(\frac{y}{|\lambda|}\right) - \alpha\right] \le |\lambda|\left[f\left(\frac{y}{|\lambda|}\right) - f\left(\frac{y}{|\lambda|}\right) + p\left(\left(\frac{y}{|\lambda|}\right) - x\right)\right]$$

$$= p(y + \lambda x).$$

Thus $g(z) \le p(z)$ for all $z \in \mathscr{M} + \mathbf{R}x$.

Evidently the same reasoning can be applied to any linear extension F of f satisfying $F \le p$ on its domain, and it shows that the domain of a maximal linear extension satisfying $F \le p$ must be the whole space $\mathscr{X}$. What remains to be shown is that such a maximal extension exists. But the family $\mathscr{F}$ of all linear extensions F of f satisfying $F \le p$ is partially ordered by inclusion (maps from subspaces of $\mathscr{X}$ to $\mathbf{R}$ being regarded as subspaces of $\mathscr{X} \times \mathbf{R}$). Since the union of an increasing family of subspaces of $\mathscr{X}$ is again a subspace, one easily sees that the union of a linearly ordered subfamily of $\mathscr{F}$ lies in $\mathscr{F}$. The proof is therefore completed by invoking Zorn's lemma. □

If p is a seminorm and f: $\mathscr{X} \to \mathbf{R}$ is linear, the inequality $f \le p$ is equivalent to the inequality $|f| \le p$, because $|f(x)| = \pm f(x) = f(\pm x)$ and $p(-x) = p(x)$. In this situation the Hahn–Banach theorem also applies to complex linear functionals:

(5.6) The Complex Hahn–Banach Theorem. Let $\mathscr{X}$ be a complex vector space, p a seminorm on $\mathscr{X}$, $\mathscr{M}$ a subspace of $\mathscr{X}$, and f a complex linear functional on $\mathscr{M}$ such that $|f(x)| \le p(x)$ for $x \in \mathscr{M}$. Then there exists a linear functional F on $\mathscr{X}$ such that $|F(x)| \le p(x)$ for all $x \in \mathscr{X}$ and $F|\mathscr{M} = f$.

Proof. Let $u = \operatorname{Re} f$. By Theorem (5.5) there is a real linear extension U of u to $\mathscr{X}$ such that $|U(x)| \le p(x)$ for all $x \in \mathscr{X}$. Let $F(x) = U(x) - iU(ix)$ as

in Proposition (5.4). Then F is a complex linear extension of f, and as in the proof of Proposition (5.4), if $\alpha = \operatorname{sgn} F(x)$ we have $|F(x)| = \alpha F(x) = U(\alpha x) \le p(\alpha x) = p(x)$. □

From now on until Section 5 all of our results apply equally to real or complex vector spaces, but for the sake of definiteness we shall assume that the scalar field is $\mathbf{C}$. The principal applications of the Hahn–Banach theorem to normed vector spaces are summarized in the following theorem.

(5.7) Theorem. Let $\mathscr{X}$ be a normed vector space.

(a) If $x \ne 0 \in \mathscr{X}$, there exists $f \in \mathscr{X}^*$ such that $\|f\| = 1$ and $f(x) = \|x\|$.

(b) The bounded linear functionals on $\mathscr{X}$ separate points.

(c) If $\mathscr{M}$ is a closed subspace of $\mathscr{X}$ and $x \in \mathscr{X} \setminus \mathscr{M}$, there exists $f \in \mathscr{X}^*$ such that $f(x) \ne 0$ and $f = 0$ on $\mathscr{M}$; in fact, if $\delta = \inf\{\|x - y\|: y \in \mathscr{M}\}$, we can take $\|f\| = 1$ and $f(x) = \delta$.

(d) If $x \in \mathscr{X}$, define $\hat{x}$: $\mathscr{X}^* \to \mathbf{C}$ by $\hat{x}(f) = f(x)$. Then the map $x \to \hat{x}$ is a linear isometry from $\mathscr{X}$ into $\mathscr{X}^{**}$ (the dual of $\mathscr{X}^*$).

Proof. (a) Apply the Hahn-Banach theorem with $p =$ the norm on $\mathscr{X}$, $\mathscr{M} = \mathbf{C}x$, and f defined on $\mathscr{M}$ by $f(\lambda x) = \lambda\|x\|$.

(b) If $x \ne y$, by (a) there exists $f \in \mathscr{X}^*$ with $f(x - y) \ne 0$, that is, $f(x) \ne f(y)$.

(c) Define f on $\mathscr{M} + \mathbf{C}x$ by $f(y + \lambda x) = \lambda\delta$ ($y \in \mathscr{M}, \lambda \in \mathbf{C}$). Since

$$\|y + \lambda x\| = |\lambda|\|\lambda^{-1}y + x\| \ge |\lambda|\delta,$$

the Hahn–Banach theorem applies.

(d) Obviously $\hat{x}$ is a linear functional on $\mathscr{X}^*$ and the map $x \to \hat{x}$ is linear. Moreover, $|\hat{x}(f)| = |f(x)| \le \|f\|\|x\|$, so $\|\hat{x}\| \le \|x\|$. On the other hand, (a) implies that $\|\hat{x}\| \ge \|x\|$. □

With notation as in Theorem (5.7d), let $\hat{\mathscr{X}} = \{\hat{x}: x \in \mathscr{X}\}$. Since $\mathscr{X}^{**}$ is always complete, the closure $\bar{\hat{\mathscr{X}}}$ of $\hat{\mathscr{X}}$ in $\mathscr{X}^{**}$ is a Banach space, and the map $x \to \hat{x}$ embeds $\mathscr{X}$ as a dense subspace of $\bar{\hat{\mathscr{X}}}$. $\bar{\hat{\mathscr{X}}}$ is called the **completion** of $\mathscr{X}$. In particular, if $\mathscr{X}$ is itself a Banach space then $\bar{\hat{\mathscr{X}}} = \hat{\mathscr{X}}$.

If $\mathscr{X}$ is finite dimensional, then of course $\hat{\mathscr{X}} = \mathscr{X}^{**}$, since these spaces have the same dimension. For infinite-dimensional Banach spaces it may or may not happen that $\hat{\mathscr{X}} = \mathscr{X}^{**}$; if it does, $\mathscr{X}$ is called **reflexive**. The examples of Banach spaces we have examined so far are not reflexive except in trivial cases where they turn out to be finite-dimensional. We shall prove some cases of this assertion and present examples of reflexive Banach spaces in later sections.

EXERCISES

21. A linear functional f on a normed vector space $\mathscr{X}$ is bounded iff $f^{-1}(\{0\})$ is closed. (Use Exercise 11b.)

22. Let $\mathscr{X}$ be a normed vector space.

(a) If $\mathscr{M}$ is a closed subspace and $x \in \mathscr{X} \setminus \mathscr{M}$ then $\mathscr{M} + \mathbf{C}x$ is closed. [Use Theorem (5.7c).]

(b) Every finite-dimensional subspace of $\mathscr{X}$ is closed.

23. Let $\mathscr{X}$ be an infinite-dimensional normed vector space.

(a) There is a sequence $\{x_j\}$ in $\mathscr{X}$ such that $\|x_j\| = 1$ for all j and $\|x_j - x_k\| \geq \frac{1}{2}$ for $j \neq k$. (Construct x_j inductively, using Exercises 11b and 22.)

(b) $\mathscr{X}$ is not locally compact.

24. If $\mathscr{M}$ is a finite-dimensional subspace of a normed vector space $\mathscr{X}$, there is a closed subspace $\mathscr{N}$ such that $\mathscr{M} \cap \mathscr{N} = \{0\}$ and $\mathscr{M} + \mathscr{N} = \mathscr{X}$.

25. If $\mathscr{X}$ and $\mathscr{Y}$ are normed vector spaces, define α: $\mathscr{X}^* \times \mathscr{Y}^* \to (\mathscr{X} \times \mathscr{Y})^*$ by $\alpha(f, g)(x, y) = f(x) + g(y)$. Then α is an isomorphism which is isometric if we use the norm $\|(x, y)\| = \max(\|x\|, \|y\|)$ on $\mathscr{X} \times \mathscr{Y}$ and the corresponding operator norm on $(\mathscr{X} \times \mathscr{Y})^*$, and the norm $\|(f, g)\| = \|f\| + \|g\|$ on $\mathscr{X}^* \times \mathscr{Y}^*$.

26. Suppose $\mathscr{X}$ and $\mathscr{Y}$ are normed vector spaces and $T \in L(\mathscr{X}, \mathscr{Y})$.

(a) Define T^*: $\mathscr{Y}^* \to \mathscr{X}^*$ by $T^*f = f \circ T$. Then $T^* \in L(\mathscr{Y}^*, \mathscr{X}^*)$ and $\|T^*\| = \|T\|$. T^* is called the **adjoint** of T.

(b) Applying the construction in (a) twice one obtains $T^{**} \in L(\mathscr{X}^{**}, \mathscr{Y}^{**})$. If $\mathscr{X}, \mathscr{Y}$ are identified with their natural images in $\mathscr{X}^{**}, \mathscr{Y}^{**}$, then $T^{**}|\mathscr{X} = T$.

(c) T^* is injective iff the range of T is dense in $\mathscr{Y}$.

(d) If the range of T^* is dense in $\mathscr{X}^*$, then T is injective; the converse is true if $\mathscr{X}$ is a reflexive Banach space.

27. Suppose $\mathscr{X}$ is a Banach space. If $\mathscr{M}$ is a closed subspace of $\mathscr{X}$ and $\mathscr{N}$ is a closed subspace of $\mathscr{X}^*$, let $\mathscr{M}^0 = \{f \in \mathscr{X}^*: f|\mathscr{M} = 0\}$ and $\mathscr{N}^\perp = \{x \in \mathscr{X}: f(x) = 0 \text{ for all } f \in \mathscr{N}\}$. (Thus, if we identify $\mathscr{X}$ with its image in $\mathscr{X}^{**}$, $\mathscr{N}^\perp = \mathscr{N}^0 \cap \mathscr{X}$.)

(a) $\mathscr{M}^0$ and $\mathscr{N}^\perp$ are closed subspaces of $\mathscr{X}^*$ and $\mathscr{X}$ respectively.

(b) $(\mathscr{M}^0)^\perp = \mathscr{M}$ and $(\mathscr{N}^\perp)^0 \supset \mathscr{N}$. If $\mathscr{X}$ is reflexive, $(\mathscr{N}^\perp)^0 = \mathscr{N}$.

(c) Let π: $\mathscr{X} \to \mathscr{X}/\mathscr{M}$ be the canonical projection, and define α: $(\mathscr{X}/\mathscr{M})^* \to \mathscr{X}^*$ by $\alpha(f) = f \circ \pi$. Then α is an isometric isomorphism from $(\mathscr{X}/\mathscr{M})^*$ onto $\mathscr{M}^0$, where $\mathscr{X}/\mathscr{M}$ has the quotient norm (cf. Exercise 11).

(d) Define β: $\mathscr{X}^* \to \mathscr{M}^*$ by $\beta(f) = (f|\mathscr{M})$. By Exercise 14, β induces a map $\bar{\beta}$: $\mathscr{X}^*/\mathscr{M}^0 \to \mathscr{M}^*$. This $\bar{\beta}$ is an isometric isomorphism.

28. Suppose $\mathscr{X}$ is a Banach space.

(a) Let $\hat{\mathscr{X}}, (\mathscr{X}^*)\hat{}$ be the natural images of $\mathscr{X}$, $\mathscr{X}^*$ in $\mathscr{X}^{**}$, $\mathscr{X}^{***}$, and let $\hat{\mathscr{X}}^0 = \{F \in \mathscr{X}^{***}: F|\hat{\mathscr{X}} = 0\}$. Then $(\mathscr{X}^*)\hat{} \cap \hat{\mathscr{X}}^0 = \{0\}$ and $(\mathscr{X}^*)\hat{} + \hat{\mathscr{X}}^0 = \mathscr{X}^{***}$.

(b) $\mathscr{X}$ is reflexive iff $\mathscr{X}^*$ is reflexive.

29. If $\mathscr{X}$ is a Banach space and $\mathscr{X}^*$ is separable, then $\mathscr{X}$ is separable. (The converse is false.) [Let $\{f_n\}_1^\infty$ be a countable dense subset of $\mathscr{X}^*$. For each n choose $x_n \in \mathscr{X}$ with $\|x_n\| = 1$ and $|f_n(x_n)| \geq \frac{1}{2}\|f_n\|$. Then the linear combinations of $\{x_n\}_1^\infty$ are dense in $\mathscr{X}$.]

30. (Krein's Extension Theorem) Let $\mathscr{X}$ be a real vector space and P a subset of $\mathscr{X}$ such that if $x, y \in P$ and $\lambda \geq 0$ then $x + y \in P$ and $\lambda x \in P$. (Example: if $\mathscr{X}$ is a space of real-valued functions, P can be the set of nonnegative functions in $\mathscr{X}$.) Define a partial ordering on $\mathscr{X}$ by declaring that $x \leq y$ iff $y - x \in P$, and call a linear functional f on $\mathscr{X}$ P-positive if $f(x) \geq 0$ for $x \in P$. Let $\mathscr{M}$ be a subspace of $\mathscr{X}$ such that for each $x \in \mathscr{X}$ there exists $y \in \mathscr{M}$ with $x \leq y$. If f is a $(P \cap \mathscr{M})$-positive linear functional on $\mathscr{M}$, there is a P-positive linear functional F on $\mathscr{X}$ such that $F|\mathscr{M} = f$. [Consider $p(x) = \inf\{f(y): y \in \mathscr{M} \text{ and } x \leq y\}$.]

3. THE BAIRE CATEGORY THEOREM AND ITS CONSEQUENCES

In this section we present an important theorem about complete metric spaces and use it to obtain some fundamental results concerning linear maps between Banach spaces.

(5.8) The Baire Category Theorem. Let X be a complete metric space.

(a) If $\{U_n\}_1^\infty$ is a sequence of open dense subsets of X, then $\bigcap_1^\infty U_n$ is dense in X.

(b) X is not a countable union of nowhere dense sets.

Proof. (a) We must show that if W is a nonempty open set in X, then $(\bigcap_1^\infty U_n) \cap W \neq \varnothing$. Since $U_1 \cap W$ is open and nonempty, it contains a ball $B(r_1, x_1)$, and we can assume that $0 < r_1 < 1$. We choose sequences $\{x_n\}$ and $\{r_n\}$ inductively as follows: having chosen x_j and r_j for $j < n$, we observe that $U_n \cap B(r_{n-1}, x_{n-1})$ is open and nonempty, so we can choose x_n, r_n so that $0 < r_n < 2^{-n}$ and $\overline{B(r_n, x_n)} \subset U_n \cap B(r_{n-1}, x_{n-1})$. Then if $n, m \geq N$, we see that $x_n, x_m \in B(r_N, x_N)$, and since $r_n \to 0$, the sequence $\{x_n\}$ is Cauchy. As X is complete, $x = \lim x_n$ exists. Since $x_n \in B(r_N, x_N)$ for $n \geq N$ we have

$$x \in \overline{B(r_N, x_N)} \subset U_N \cap B(r_1, x_1) \subset U_N \cap W$$

for all N, and the proof is complete.

(b) If $\{E_n\}$ is a sequence of nowhere dense sets, then $\{(\overline{E}_n)^c\}$ is a sequence of open dense sets. Since $\bigcap_1^\infty (\overline{E}_n)^c \neq \emptyset$, we have $\bigcup_1^\infty E_n \subset \bigcup_1^\infty \overline{E}_n \neq X$. □

We remark that since the conclusions of the Baire category theorem are purely topological, it suffices for X to be homeomorphic to a complete metric space. For example, the theorem applies to $X = (0,1)$, which is not complete with the usual metric but is homeomorphic to **R**.

The name of this theorem comes from Baire's terminology for sets: if X is a topological space, a set $E \subset X$ is **of the first category**, according to Baire, if E is a countable union of nowhere dense sets; otherwise E is **of the second category**. Thus Baire's theorem asserts that every complete metric space is of the second category in itself. A more modern and more descriptive synonym for "of the first category" is **meager**. The complement of a meager set is called **residual**.

The Baire category theorem is often used to prove existence results: one shows that objects having a certain property exist by showing that the set of objects *not* having the property (within a suitable complete metric space) is meager. For example, one can prove the existence of nowhere differentiable continuous functions in this way: see Exercise 46.

We turn to the applications of the Baire category theorem to the theory of linear maps. Some terminology: if X and Y are topological spaces, a map f: $X \to Y$ is called **open** if $f(U)$ is open in Y whenever U is open in X. If X and Y are metric spaces, this amounts to requiring that if B is a ball centered at $x \in X$, then $f(B)$ contains a ball centered at $f(x)$. Finally, if X and Y are normed linear spaces and f is linear, so that f commutes with translations and dilations, f is open iff $f(B)$ contains a ball centered at 0 in Y, where B is the ball of radius 1 about 0 in X.

(5.9) The Open Mapping Theorem. Let $\mathscr{X}$ and $\mathscr{Y}$ be Banach spaces. If $T \in L(\mathscr{X}, \mathscr{Y})$ is surjective, then T is open.

Proof. Let B_r denote the (open) ball of radius r about 0 in $\mathscr{X}$. By the preceding remarks it will suffice to show that $T(B_1)$ contains a ball about 0 in $\mathscr{Y}$. Since $\mathscr{X} = \bigcup_1^\infty B_n$ and T is surjective, we have $\mathscr{Y} = \bigcup_1^\infty T(B_n)$. But $\mathscr{Y}$ is complete and the map $y \to ny$ is a homeomorphism of $\mathscr{Y}$ which maps $T(B_1)$ to $T(B_n)$, so Baire's theorem implies that $T(B_1)$ cannot be nowhere dense. That is, there exist $y_0 \in \mathscr{Y}$ and $r > 0$ such that the ball $B(4r, y_0)$ is contained in $\overline{T(B_1)}$. Pick $y_1 = Tx_1 \in T(B_1)$ such that $\|y_1 - y_0\| < 2r$: then $B(2r, y_1) \subset \overline{T(B_1)}$, so if $\|y\| < 2r$,

$$y = Tx_1 + (y - y_1) \in \overline{T(x_1 + B_1)} \subset \overline{T(B_2)}.$$

Dividing both sides by 2, we conclude that there exists $r > 0$ such that if $\|y\| < r$ then $y \in \overline{T(B_1)}$. If we could replace $\overline{T(B_1)}$ by $T(B_1)$, perhaps

shrinking r at the same time, the proof would be complete; we now proceed to accomplish this.

From dilation invariance it follows that if $\|y\| < r2^{-n}$ then $y \in \overline{T(B_{2^{-n}})}$. Suppose $\|y\| < r/2$: we can find $x_1 \in B_{1/2}$ such that $\|y - Tx_1\| < r/4$, and proceeding inductively we can find $x_n \in B_{2^{-n}}$ such that $\|y - \sum_1^n Tx_j\| < r2^{-n-1}$. Since $\mathscr{X}$ is complete, by Theorem (5.1) the series $\sum_1^\infty x_n$ converges, say to x. But then $\|x\| < \sum_1^\infty 2^{-n} = 1$ and $y = Tx$. In other words, $T(B_1)$ contains all y with $\|y\| < r/2$, so we are done. □

(5.10) Corollary. If $\mathscr{X}$ and $\mathscr{Y}$ are Banach spaces and $T \in L(\mathscr{X}, \mathscr{Y})$ is bijective, then T is an isomorphism; that is, $T^{-1} \in L(\mathscr{Y}, \mathscr{X})$.

Proof. If T is bijective, continuity of T^{-1} is equivalent to the openness of T. □

For the next result we need some more terminology. If $\mathscr{X}$ and $\mathscr{Y}$ are normed vector spaces and T is a linear map from $\mathscr{X}$ to $\mathscr{Y}$, we define the **graph** of T to be

$$\Gamma(T) = \{(x, y) \in \mathscr{X} \times \mathscr{Y} : y = Tx\},$$

which is a subspace of $\mathscr{X} \times \mathscr{Y}$. [From a strict set-theoretic point of view, of course, T and $\Gamma(T)$ are identical; the distinction is a psychological one.] We say that T is **closed** if $\Gamma(T)$ is a closed subspace of $\mathscr{X} \times \mathscr{Y}$. Clearly, if T is continuous then T is closed, and if $\mathscr{X}$ and $\mathscr{Y}$ are complete the converse is also true:

(5.11) The Closed Graph Theorem. If $\mathscr{X}$ and $\mathscr{Y}$ are Banach spaces and T: $\mathscr{X} \to \mathscr{Y}$ is a closed linear map, then T is bounded.

Proof. Let π_1, π_2 be the projections of $\Gamma(T)$ onto $\mathscr{X}$ and $\mathscr{Y}$: $\pi_1(x, Tx) = x$ and $\pi_2(x, Tx) = Tx$. Obviously $\pi_1 \in L(\Gamma(T), \mathscr{X})$ and $\pi_2 \in L(\Gamma(T), \mathscr{Y})$. Since $\mathscr{X}$ and $\mathscr{Y}$ are complete, so is $\mathscr{X} \times \mathscr{Y}$, and hence so is $\Gamma(T)$ since T is closed. The map π_1 is a bijection from $\Gamma(T)$ to $\mathscr{X}$, so by Corollary (5.10), π_1^{-1} is bounded. But then $T = \pi_2 \circ \pi_1^{-1}$ is bounded. □

We remark that T: $\mathscr{X} \to \mathscr{Y}$ is continuous iff whenever $x_n \to x$ then $Tx_n \to Tx$, and T is closed iff whenever $x_n \to x$ and $Tx_n \to y$ then $y = Tx$. Thus the significance of the closed graph theorem is that in verifying that $Tx_n \to Tx$ when $x_n \to x$, we may assume that Tx_n converges to *something*, and we need only show that the limit is the right thing. This frequently saves a lot of trouble.

The completeness of $\mathscr{X}$ and $\mathscr{Y}$ was used in a crucial way in proving the open mapping theorem and hence also in proving the closed graph theorem. In fact, the conclusions of both of these theorems may fail if either $\mathscr{X}$ or $\mathscr{Y}$ is incomplete: see Exercises 33–35.

Our final result in this section allows one to deduce uniform estimates from pointwise estimates in certain situations.

(5.12) The Uniform Boundedness Principle. Suppose $\mathscr{X}$ and $\mathscr{Y}$ are normed vector spaces and $\mathscr{A}$ is a subset of $L(\mathscr{X}, \mathscr{Y})$.

(a) If $\sup_{T \in \mathscr{A}} \|Tx\| < \infty$ for all x in a nonmeager subset of $\mathscr{X}$, then $\sup_{T \in \mathscr{A}} \|T\| < \infty$.

(b) If $\mathscr{X}$ is a Banach space and $\sup_{T \in \mathscr{A}} \|Tx\| < \infty$ for all $x \in \mathscr{X}$, then $\sup_{T \in \mathscr{A}} \|T\| < \infty$.

Proof. Let

$$E_n = \{x \in \mathscr{X} : \sup_{T \in \mathscr{A}} \|Tx\| \leq n\} = \bigcap_{T \in \mathscr{A}} \{x \in \mathscr{X} : \|Tx\| \leq n\}.$$

Then the E_n's are closed, so under the hypothesis of (a) some E_n must contain a nontrivial closed ball $\overline{B(r, x_0)}$. But then $E_{2n} \supset \overline{B(r, 0)}$, for if $\|x\| \leq r$ then $x - x_0 \in E_n$ and hence

$$\|Tx\| \leq \|T(x - x_0)\| + \|Tx_0\| \leq 2n.$$

In other words, $\|Tx\| \leq 2n$ whenever $T \in \mathscr{A}$ and $\|x\| \leq r$, so $\sup_{T \in \mathscr{A}} \|T\| \leq 2n/r$. (b) follows from (a) by the Baire category theorem. □

EXERCISES

31. There exist residual subsets of $\mathbf{R}$ which have Lebesgue measure zero.

32. The conclusion of the Baire category theorem is true if X is assumed to be an LCH space rather than a complete metric space. [The proof is similar; the substitute for completeness is Proposition (4.21).]

33. Let $\mathscr{Y} = L^1(\mu)$ where μ is counting measure on $\mathbf{N}$, and let $\mathscr{X} = \{f \in \mathscr{Y} : \sum_1^\infty n|f(n)| < \infty\}$ (with the L^1 norm).

(a) $\mathscr{X}$ is a proper dense subspace of $\mathscr{Y}$; hence $\mathscr{X}$ is not complete.

(b) Define $T: \mathscr{X} \to \mathscr{Y}$ by $Tf(n) = nf(n)$. Then T is closed but not bounded.

(c) Let $S = T^{-1}$. Then $S: \mathscr{Y} \to \mathscr{X}$ is bounded and surjective but not open.

34. Let $\mathscr{Y} = C([0,1])$ and $\mathscr{X} = C^1([0,1])$, both equipped with the uniform norm.

(a) $\mathscr{X}$ is not complete.

(b) The map $(d/dx): \mathscr{X} \to \mathscr{Y}$ is closed (see Exercise 7) but not bounded.

35. Let $\mathscr{X}, \mathscr{Y}$ be Banach spaces and let $S: \mathscr{X} \to \mathscr{Y}$ be an unbounded linear map (for the existence of which, see Section 6). Let $\Gamma(S)$ be the graph of S, a subspace of $\mathscr{X} \times \mathscr{Y}$.

(a) $\Gamma(S)$ is not complete.

(b) Define $T: \mathscr{X} \to \Gamma(S)$ by $Tx = (x, Sx)$. Then T is closed but unbounded.

(c) $T^{-1}: \Gamma(S) \to \mathscr{X}$ is bounded and surjective but not open.

36. Let $\| \|_1$ and $\| \|_2$ be norms on the vector space $\mathscr{X}$ such that $\| \|_1 \leq \| \|_2$ and $\mathscr{X}$ is complete with respect to both norms. Then $\| \|_1$ and $\| \|_2$ are equivalent.

37. There is no slowest rate of decay for the terms of an absolutely convergent series; that is, there is no sequence $\{a_n\}$ of positive numbers such that $\sum a_n|c_n| < \infty$ iff $\{c_n\}$ is bounded. [The set of bounded sequences is the space $B(\mathbf{N})$ of bounded functions on $\mathbf{N}$, and the space of absolutely summable sequences is $L^1(\mu)$ where μ is counting measure on $\mathbf{N}$. If such an $\{a_n\}$ exists, consider $T: B(\mathbf{N}) \to L^1(\mu)$ defined by $Tf(n) = a_n f(n)$. $\{f: f(n) = 0$ for all but finitely many $n\}$ is dense in $L^1(\mu)$ but not in $B(\mathbf{N})$.]

38. With reference to Exercises 7 and 8, show that the inclusion map of $C^k([0,1])$ into $L^1_k([0,1])$ is continuous (a) by using the closed graph theorem, and (b) by direct calculation. (This is to illustrate the use of the closed graph theorem as a labor-saving device.)

39. Let $\mathscr{X}$ and $\mathscr{Y}$ be Banach spaces, $T \in L(\mathscr{X}, \mathscr{Y})$, $\mathscr{N}(T) = \{x: Tx = 0\}$, and $\mathscr{M} = \text{range}(T)$. Then $X/\mathscr{N}(T)$ is isomorphic to $\mathscr{M}$ iff $\mathscr{M}$ is closed. (Cf. Exercise 11.)

40. Let $\mathscr{X}$ be a separable Banach space and let μ be counting measure on $\mathbf{N}$. Suppose $\{x_n\}$ is a countable dense subset of the unit ball of $\mathscr{X}$, and define $T: L^1(\mu) \to \mathscr{X}$ by $Tf = \sum_1^\infty f(n)x_n$.

(a) T is bounded.

(b) T is surjective.

(c) $\mathscr{X}$ is isomorphic to a quotient space of $L^1(\mu)$. (Use Exercise 39.)

41. Let $\mathscr{X}$ and $\mathscr{Y}$ be Banach spaces. If $T: \mathscr{X} \to \mathscr{Y}$ is a linear map such that $f \circ T \in \mathscr{X}^*$ for every $f \in \mathscr{Y}^*$, then T is bounded.

42. Let $\mathscr{X}$ and $\mathscr{Y}$ be Banach spaces, and let $\{T_n\}$ be a sequence in $L(\mathscr{X}, \mathscr{Y})$ such that $\lim T_n x$ exists for every $x \in \mathscr{X}$. If $Tx = \lim T_n x$, then $T \in L(\mathscr{X}, \mathscr{Y})$.

43. Let $\mathscr{X}$, $\mathscr{Y}$, $\mathscr{Z}$ be Banach spaces and let $B: \mathscr{X} \times \mathscr{Y} \to \mathscr{Z}$ be a separately continuous bilinear map, that is, $B(x, \cdot) \in L(\mathscr{Y}, \mathscr{Z})$ for each $x \in \mathscr{X}$ and $B(\cdot, y) \in L(\mathscr{X}, \mathscr{Z})$ for each $y \in \mathscr{Y}$. Then B is jointly continuous, that is, continuous from $\mathscr{X} \times \mathscr{Y}$ to $\mathscr{Z}$. [Reduce the problem to proving that $\|B(x, y)\| \leq C\|x\| \|y\|$ for some $C > 0$.]

44. (The Principle of Condensation of Singularities) Let $\mathscr{X}$ and $\mathscr{Y}$ be Banach spaces and $\{T_{jk}: j,k \in \mathbf{N}\} \subset L(\mathscr{X},\mathscr{Y})$. Suppose that for each k there exists $x \in \mathscr{X}$ such that $\sup\{\|T_{jk}x\|: j \in \mathbf{N}\} = \infty$. Then there is an x (indeed, a residual set of x's) such that $\sup\{\|T_{jk}x\|: j,k \in \mathbf{N}\} = \infty$.

45. Let $\mathscr{X}$ be a vector space of countably infinite dimension (that is, every element is a *finite* linear combination of members of a countably infinite linearly independent set). There is no norm on $\mathscr{X}$ with respect to which $\mathscr{X}$ is complete. (Given a norm on $\mathscr{X}$, apply Exercise 22b and the Baire category theorem.)

46. Let E_n be the set of all $f \in C([0,1])$ for which there exists $x_0 \in [0,1]$ (depending on f) such that $|f(x) - f(x_0)| \le n|x - x_0|$ for all $x \in [0,1]$.

(a) E_n is nowhere dense in $C([0,1])$. [Any real $f \in C([0,1])$ can be uniformly approximated by a piecewise linear function g whose linear pieces, finite in number, have slope $\pm 2n$. If $\|h - g\|_u$ is sufficiently small then $h \notin E_n$.]

(b) The set of nowhere differentiable functions is residual in $C([0,1])$.

4. TOPOLOGICAL VECTOR SPACES

It is frequently useful to consider topologies on vector spaces other than those defined by norms, the only crucial requirement being that the topology should be well behaved with respect to the vector operations. Precisely, a **topological vector space** is a vector space $\mathscr{X}$ over the field K ($= \mathbf{R}$ or $\mathbf{C}$) which is endowed with a topology such that the maps $(x,y) \to x + y$ and $(\lambda, x) \to \lambda x$ are continuous from $\mathscr{X} \times \mathscr{X}$ and $K \times \mathscr{X}$ to $\mathscr{X}$. A topological vector space is called **locally convex** if there is a base for the topology consisting of convex sets [i.e., sets A such that if $x, y \in A$ then $tx + (1-t)y \in A$ for $0 < t < 1$]. Most topological vector spaces that arise in practice are locally convex and Hausdorff.

In a topological vector space $\mathscr{X}$ the notion of Cauchy sequence or net makes sense. Indeed, if A is a directed set and $A \times A$ is directed in the standard way [$(\alpha,\beta) \le (\alpha',\beta')$ iff $\alpha \le \alpha'$ and $\beta \le \beta'$], a net $\langle x_\alpha \rangle_{\alpha \in A}$ is called **Cauchy** if the net $\langle x_\alpha - x_\beta \rangle_{(\alpha,\beta) \in A \times A}$ converges to zero. Naturally, $\mathscr{X}$ is called **complete** if every Cauchy net converges. A complete, metrizable, locally convex topological vector space is called a **Fréchet space**.

The most common way of defining locally convex topologies on vector spaces is in terms of seminorms. If $\mathscr{X}$ is a vector space and $\{p_\alpha\}_{\alpha \in A}$ is a family of seminorms on $\mathscr{X}$, the convex sets

$$(5.13) \quad U_{x\alpha\varepsilon} = \{y \in \mathscr{X}: p_\alpha(x - y) < \varepsilon\} \quad [x \in \mathscr{X}, \alpha \in A, \varepsilon \in (0,\infty)]$$

generate a topology on $\mathscr{X}$ with respect to which $\mathscr{X}$ is a locally convex topological vector space. The continuity of the vector operations is an easy consequence of Proposition (4.19) and the second part of the following proposition.

(5.14) Proposition. Let $\mathscr{X}$ have the topology generated by the sets $U_{x\alpha\varepsilon}$ in (5.13).

(a) For each $x \in \mathscr{X}$, the finite intersections of the sets $U_{x\alpha\varepsilon}$ ($\alpha \in A, \varepsilon > 0$) form a neighborhood base at x.

(b) If $\langle x_\gamma \rangle$ is a net in $\mathscr{X}$, then $x_\gamma \to x$ iff $p_\alpha(x_\gamma - x) \to 0$ for all α.

Proof. (a) If $x \in \bigcap_1^k U_{x_j\alpha_j\varepsilon_j}$, let $\delta_j = \varepsilon_j - p_{\alpha_j}(x - x_j)$. By the triangle inequality, we have $x \in \bigcap_1^k U_{x\alpha_j\delta_j} \subset \bigcap_1^k U_{x_j\alpha_j\varepsilon_j}$. Thus the assertion follows from Proposition (4.4).

(b) In view of (a), it suffices to observe that $p_\alpha(x_\gamma - x) \to 0$ iff $\langle x_\gamma \rangle$ is eventually in $U_{x\alpha\varepsilon}$ for every $\varepsilon > 0$. □

In this context there is an analog of Proposition (5.2):

(5.15) Proposition. Suppose $\mathscr{X}$ and $\mathscr{Y}$ are vector spaces with topologies defined, respectively, by the families $\{p_\alpha\}_{\alpha \in A}$ and $\{q_\beta\}_{\beta \in B}$ of seminorms, and $T: \mathscr{X} \to \mathscr{Y}$ is a linear map. Then T is continuous iff for each $\beta \in B$ there exist $\alpha_1, \dots, \alpha_k \in A$ and $C > 0$ such that $q_\beta(Tx) \le C\sum_1^k p_{\alpha_j}(x)$.

Proof. If the latter condition holds and $\langle x_\gamma \rangle$ is a net converging to $x \in \mathscr{X}$, by Proposition (5.14b) we have $p_\alpha(x_\gamma - x) \to 0$ for all α, hence $q_\beta(Tx_\gamma - Tx) \to 0$ for all β, hence $Tx_\gamma \to Tx$. By Proposition (4.19), T is continuous. Conversely, if T is continuous, for every $\beta \in B$ there is a neighborhood U of 0 in $\mathscr{X}$ such that $q_\beta(Tx) < 1$ for $x \in U$. By Proposition (5.14a) we may assume that $U = \bigcap_1^k U_{x\alpha_j\varepsilon_j}$. Let $\varepsilon = \min(\varepsilon_1, \dots, \varepsilon_k)$: then $q_\beta(Tx) < 1$ whenever $p_{\alpha_j}(x) < \varepsilon$ for all j. Now, given $x \in \mathscr{X}$, there are two possibilities. If $p_{\alpha_j}(x) = 0$ for all j, then $p_{\alpha_j}(rx) = 0$ for all j and all $r > 0$, hence $rq_\beta(Tx) = q_\beta(T(rx)) < 1$ for all $r > 0$, hence $q_\beta(Tx) = 0$. On the other hand, if $p_{\alpha_j}(x) > 0$ for some j, let $y = \varepsilon x / \sum_1^k p_{\alpha_j}(x)$. Then $p_{\alpha_j}(y) < \varepsilon$ for all j, so

$$q_\beta(Tx) = \sum_1^k \varepsilon^{-1} p_{\alpha_j}(x) q_\beta(Ty) \le \varepsilon^{-1} \sum_1^k p_{\alpha_j}(x). \qquad \square$$

(5.16) Proposition. Let $\mathscr{X}$ be a vector space with the topology defined by the family $\{p_\alpha\}_{\alpha \in A}$ of seminorms.

(a) $\mathscr{X}$ is Hausdorff iff for each $x \neq 0$ there exists $\alpha \in A$ such that $p_\alpha(x) \neq 0$.

(b) If $\mathscr{X}$ is Hausdorff and A is countable, then $\mathscr{X}$ is metrizable.

The proof is left to the reader (Exercise 47).

If $\mathscr{X}$ has the topology defined by the seminorms $\{p_\alpha\}_{\alpha \in A}$, by Proposition (5.15) a linear functional f on $\mathscr{X}$ is continuous iff $|f(x)| \le C\sum_1^k p_{\alpha_j}(x)$ for

some $C > 0$ and $\alpha_1, \dots, \alpha_k \in A$. Since a finite sum of seminorms is again a seminorm, the Hahn–Banach theorem guarantees the existence of lots of continuous linear functionals on $\mathscr{X}$—enough to separate points, if $\mathscr{X}$ is Hausdorff. The set of all such functionals is denoted, as before, by $\mathscr{X}^*$. There are various ways of making $\mathscr{X}^*$ into a topological vector space, but we shall not consider this question here.

We now present a list of some important examples of topologies defined by families of seminorms.

(i) Let X be an LCH space. On $\mathbf{C}^X$, the topology of uniform convergence on compact sets is defined by the seminorms $p_K(f) = \sup_{x \in K}|f(x)|$ as K ranges over all compact subsets of X. $\mathbf{C}^X$ is easily seen to be complete, and hence by Proposition (4.38), so is $C(X)$. If X is σ-compact and $\{U_n\}$ are as in Proposition (4.40), the topology of uniform convergence on compact sets is defined by the seminorms $p_n(f) = \sup_{x \in \bar{U}_n}|f(x)|$, so by Proposition (5.16) $C(X)$ is a Fréchet space.

(ii) The space $L^1_{loc}(\mathbf{R}^n)$, defined in Section 3.4, is a Fréchet space with the topology defined by the seminorms $p_k(f) = \int_{|x|<k}|f(x)|\,dx$. [There is an obvious analog of this construction which yields a locally convex space $L^1_{loc}(X,\mu)$ where X is any LCH space and μ is a Borel measure on X which is finite on compact sets.]

(iii) One of the most important linear operators in analysis is differentiation. One often wishes to study d/dx (or more complicated operators constructed from it) acting on various spaces of functions. Unfortunately, it is virtually impossible to define norms on most infinite-dimensional function spaces so that d/dx becomes a bounded operator. One precise result along these lines is the following: There is no norm on the space $C^\infty([0,1])$ of infinitely differentiable functions on $[0,1]$ with respect to which d/dx is bounded. [Proof: let $f_\lambda(x) = e^{\lambda x}$. Then $(d/dx)f_\lambda = \lambda f_\lambda$, so $\|d/dx\| \geq |\lambda|$ for all λ no matter what norm is used on $C^\infty([0,1])$.]

In view of this difficulty, three courses of action are available. First, one can consider differentiation as a unbounded operator from $\mathscr{X}$ to $\mathscr{Y}$ where $\mathscr{Y}$ is a suitable Banach space and $\mathscr{X}$ is a dense subspace of $\mathscr{Y}$, as in Exercise 34. Second, one can consider differentiation as a bounded linear map from one Banach space $\mathscr{X}$ to a different one $\mathscr{Y}$, such as $\mathscr{X} = C^k([0,1])$ and $\mathscr{Y} = C^{k-1}([0,1])$ in Exercise 7. Finally, one can consider differentiation as a continuous operator on a locally convex space $\mathscr{X}$ whose topology is not given by a norm. All of these points of view have their advantages, but it is the last one that concerns us here. It is easy to construct families of seminorms on spaces of smooth functions such that differentiation becomes continuous almost by definition. For example, the seminorms $p_k(f) = \sup_{0 \leq x \leq 1}|f^{(k)}(x)|$ $(k = 0, 1, 2, \dots)$ make $C^\infty([0,1])$ into a Fréchet space (the completeness follows as in Exercise 7), and d/dx is continuous by Proposition (5.15) since $p_k(f') = p_{k+1}(f)$. Other examples are considered in Exercise 48 and in Chapter 8.

(iv) Let $\mathscr{X}$ be a vector space, $\mathscr{Y}$ a normed vector space, and $\{T_\alpha\}_{\alpha \in A}$ a collection of linear maps from $\mathscr{X}$ to $\mathscr{Y}$. Then the weak topology $\mathscr{T}$ generated by $\{T_\alpha\}$ makes $\mathscr{X}$ into a locally convex topological vector space. Indeed, $\mathscr{T}$ is just the topology $\mathscr{T}'$ defined by the seminorms $p_\alpha(x) = \|T_\alpha x\|$. [$\mathscr{T}$ is generated by sets of the form $\{x\colon \|T_\alpha x - y_0\| < \varepsilon\}$ $(y_0 \in \mathscr{Y})$, while $\mathscr{T}'$ is generated by the sets of the form $\{x\colon \|T_\alpha x - T_\alpha x_0\| < \varepsilon\}$ $(x_0 \in \mathscr{X})$. If the T_α's are surjective, these are obviously the same; the general case is left as an exercise.] The following are among the most important examples of this construction.

Let $\mathscr{X}$ be a normed vector space. The weak topology generated by $\mathscr{X}^*$ is known simply as the **weak topology** on $\mathscr{X}$, and convergence with respect to this topology is called **weak convergence**. Thus, if $\langle x_\alpha \rangle$ is a net in $\mathscr{X}$, $x_\alpha \to x$ weakly iff $f(x_\alpha) \to f(x)$ for all $f \in \mathscr{X}^*$. When $\mathscr{X}$ is infinite-dimensional, the weak topology is always weaker than the norm topology: see Exercise 50.

Let $\mathscr{X}$ be a normed vector space, $\mathscr{X}^*$ its dual. The weak topology on $\mathscr{X}^*$ as defined above is the topology generated by $\mathscr{X}^{**}$; of more interest is the topology generated by $\mathscr{X}$ (considered as a subspace of $\mathscr{X}^{**}$), which is called the **weak * topology** (read "weak star topology") on $\mathscr{X}^*$. $\mathscr{X}^*$ is a space of functions on $\mathscr{X}$, and the weak * topology is simply the topology of pointwise convergence: $f_\alpha \to f$ iff $f_\alpha(x) \to f(x)$ for all $x \in \mathscr{X}$. The weak * topology is even weaker than the weak topology; the two coincide precisely when $\mathscr{X}$ is reflexive.

Let $\mathscr{X}$ and $\mathscr{Y}$ be Banach spaces. The topology on $L(\mathscr{X},\mathscr{Y})$ generated by the evaluation maps $T \to Tx$ $(x \in \mathscr{X})$ is called the **strong operator topology** on $L(\mathscr{X},\mathscr{Y})$, while the topology generated by the linear functionals $T \to f(Tx)$ $(x \in \mathscr{X}, f \in \mathscr{Y}^*)$ is called the **weak operator topology** on $L(\mathscr{X},\mathscr{Y})$. Again, these topologies are best understood in terms of convergence: $T_\alpha \to T$ strongly iff $T_\alpha x \to Tx$ in the norm topology of $\mathscr{Y}$ for each $x \in \mathscr{X}$, while $T_\alpha \to T$ weakly iff $T_\alpha x \to Tx$ in the weak topology of $\mathscr{Y}$ for each $x \in \mathscr{X}$. Thus the strong operator topology is stronger than the weak operator topology but weaker than the norm topology on $L(\mathscr{X},\mathscr{Y})$. The following result concerning strong convergence is almost trivial but extremely useful:

(5.17) Proposition. If $T, T_n \in L(\mathscr{X},\mathscr{Y})$ $(n \in \mathbf{N})$, $\sup_n\|T_n\| < \infty$, and $\|T_n x - Tx\| \to 0$ for all x in a dense subset D of $\mathscr{X}$, then $T_n \to T$ strongly.

Proof. Let $C = \sup\{\|T\|, \|T_1\|, \|T_2\|, \dots\}$. Given $x \in \mathscr{X}$ and $\varepsilon > 0$, choose $y \in D$ such that $\|x - y\| < \varepsilon/3C$. If n is large enough so that $\|T_n y - Ty\| < \varepsilon/3$, we have

$$\|T_n x - Tx\| \leq \|T_n x - T_n y\| + \|T_n y - Ty\| + \|Ty - Tx\|$$

$$\leq 2C\|x - y\| + \varepsilon/3 < \varepsilon,$$

so that $T_n x \to Tx$. □

Our final result in this section is a compactness theorem which is one of the main reasons for the usefulness of the weak * topology on a dual space. The idea of the proof is similar to the techniques discussed in Section 4.8.

(5.18) Alaoglu's Theorem. If $\mathscr{X}$ is a normed vector space, the closed unit ball $B = \{f \in \mathscr{X}^*: \|f\| \le 1\}$ in $\mathscr{X}^*$ is compact in the weak * topology.

Proof. For each $x \in \mathscr{X}$ let $D_x = \{z \in \mathbf{C}: |z| \le \|x\|\}$, and let $D = \prod_{x \in \mathscr{X}} D_x$. Then D is compact by Tychonoff's theorem. The elements of D are precisely those complex-valued functions φ on $\mathscr{X}$ such that $|\varphi(x)| \le \|x\|$ for all x, and B consists of those elements of D which are linear. Moreover, the relative topologies that B inherits from D and from $\mathscr{X}^*$ (with the weak * topology) both coincide with the topology of pointwise convergence, so it suffices to see that B is closed in D. But this is easy: if $\langle f_\alpha \rangle$ is a net in B which converges to $f \in D$, for any $x, y \in \mathscr{X}$ and $a, b \in \mathbf{C}$ we have

$$f(ax + by) = \lim f_\alpha(ax + by) = \lim(af_\alpha(x) + bf_\alpha(y)) = af(x) + bf(y),$$

so that $f \in B$. □

Warning: Alaoglu's theorem does *not* say that $\mathscr{X}^*$ is locally compact in the weak * topology: see Exercise 50.

EXERCISES

47. Prove Proposition (5.16). [For part (b), proceed as in Exercise 4.55.]

48. The space $C^\infty(\mathbf{R})$ of all infinitely differentiable functions on $\mathbf{R}$ has a Fréchet space topology in which "$f_n \to f$" means that $f_n^{(k)} \to f^{(k)}$ uniformly on compact sets for all $k \ge 0$.

49. If $\mathscr{X}$ is a vector space, $\mathscr{Y}$ a normed vector space, $\mathscr{T}$ the topology on $\mathscr{X}$ generated by a family of linear maps $\{T_\alpha: \mathscr{X} \to \mathscr{Y}\}$, and $\mathscr{T}'$ the topology defined by the family of seminorms $\{x \to \|T_\alpha x\|\}$, then $\mathscr{T} = \mathscr{T}'$.

50. Let $\mathscr{X}$ be a normed vector space.

(a) Every weakly convergent sequence in $\mathscr{X}$, and every weak * convergent sequence in $\mathscr{X}^*$, is bounded with respect to the norm. (Use the uniform boundedness principle.)

(b) Every weakly compact subset of $\mathscr{X}$, and every weak * compact subset of $\mathscr{X}^*$, is bounded with respect to the norm.

(c) If $\mathscr{X}$ is infinite-dimensional, every nonempty weakly open set in $\mathscr{X}$, and every nonempty weak * open set in $\mathscr{X}^*$, is unbounded with respect to the norm.

51. If $\mathscr{X}$ is a separable normed vector space, the weak * topology on the closed unit ball in $\mathscr{X}^*$ is second countable and hence metrizable. (This does not say that $\mathscr{X}^*$ is weak * metrizable.)

52. A linear subspace of a normed vector space $\mathscr{X}$ is norm closed iff it is weakly closed. (This is not true if $\mathscr{X} = \mathscr{Y}^*$ and the weak topology is replaced by the weak * topology, unless $\mathscr{X}$ is reflexive. See Exercise 53.)

53. Let $\mathscr{X}$ be a Banach space and let $f_1, \dots, f_n$ be linearly independent elements of $\mathscr{X}^*$.

(a) Define $T: \mathscr{X} \to \mathbf{C}^n$ by $Tx = (f_1(x), \dots, f_n(x))$. If $\mathscr{N} = \{x: Tx = 0\}$ and $\mathscr{M}$ is the linear span of $f_1, \dots, f_n$, then $\mathscr{M} = \mathscr{N}^0$ in the notation of Exercise 27 and hence $\mathscr{M}^*$ is isomorphic to $(\mathscr{X}/\mathscr{N})^*$.

(b) If $F \in \mathscr{X}^{**}$, for any $\varepsilon > 0$ there exists $x \in \mathscr{X}$ such that $F(f_j) = f_j(x)$ for $j = 1, \dots, n$ and $\|x\| \le (1 + \varepsilon)\|F\|$. [$F|\mathscr{M}$ can be identified with an element of $(\mathscr{X}/\mathscr{N})^{**}$ and hence with an element of $\mathscr{X}/\mathscr{N}$ since the latter is finite-dimensional.]

(c) If $\mathscr{X}$ is considered as a subspace of $\mathscr{X}^{**}$, the relative topology on $\mathscr{X}$ induced by the weak * topology on $\mathscr{X}^{**}$ is the weak topology on $\mathscr{X}$.

(d) In the weak * topology on $\mathscr{X}^{**}$, $\mathscr{X}$ is dense in $\mathscr{X}^{**}$ and the closed unit ball in $\mathscr{X}$ is dense in the closed unit ball in $\mathscr{X}^{**}$.

(e) $\mathscr{X}$ is reflexive iff its closed unit ball is weakly compact.

54. Suppose $\mathscr{X}$ is a Banach space and $\{T_n\}$, $\{S_n\}$ are sequences in $L(\mathscr{X}, \mathscr{X})$.

(a) If $T_n \to T$ strongly and $S_n \to S$ strongly, then $T_nS_n \to TS$ strongly.

(b) If $T_n \to T$ strongly and $x_n \to x$ in the norm topology, then $T_nx_n \to Tx$ in the norm topology.

(Use Exercise 43.)

5. HILBERT SPACES

The most important Banach spaces, and the ones on which the most refined analysis can be done, are the Hilbert spaces, which are a direct generalization of finite-dimensional Euclidean spaces. Before defining them, we need to introduce a few concepts.

Let $\mathscr{X}$ be a complex vector space. A **scalar product** (or **inner product**) on $\mathscr{X}$ is a map $(x, y) \to \langle x, y \rangle$ from $\mathscr{X} \times \mathscr{X}$ to $\mathbf{C}$ such that:

(i) $\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle$ for all $x, y, z \in \mathscr{X}$ and $a, b \in \mathbf{C}$.

(ii) $\langle y, x \rangle = \overline{\langle x, y \rangle}$ for all $x, y \in \mathscr{X}$.

(iii) $\langle x, x \rangle \in (0, \infty)$ for all nonzero $x \in \mathscr{X}$.

We observe that (i) and (ii) imply

(i′) $\langle x, ay + bz \rangle = \bar{a}\langle x, y \rangle + \bar{b}\langle x, z \rangle$ for all $x, y, z \in \mathscr{X}$ and $a, b \in \mathbf{C}$.

[One can also define scalar products on real vector spaces: $\langle x, y \rangle$ is then real, a and b are assumed real in (i), and (ii) becomes $\langle y, x \rangle = \langle x, y \rangle$.] A complex

vector space equipped with a scalar product is called a **pre-Hilbert space**. If $\mathscr{X}$ is a pre-Hilbert space, for $x \in \mathscr{X}$ we define

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

(5.19) The Schwarz Inequality. $|\langle x, y\rangle| \le \|x\|\|y\|$ for all $x, y \in \mathscr{X}$, with equality iff x and y are linearly dependent.

Proof. If $\langle x, y\rangle = 0$, the result is obvious. If $\langle x, y\rangle \neq 0$ (and in particular $y \neq 0$), let $\alpha = \operatorname{sgn}\langle x, y\rangle$ and $z = \alpha y$. Then for $t \in \mathbf{R}$ we have

$$0 \le \langle x - tz, x - tz\rangle = \langle x, x\rangle - t\langle x, z\rangle - t\langle z, x\rangle + t^2\langle z, z\rangle$$

$$= \|x\|^2 - 2t|\langle x, y\rangle| + t^2\|y\|^2.$$

This last expression is a quadratic function of t whose absolute minimum occurs at $t = \|y\|^{-2}|\langle x, y\rangle|$. Substituting this value for t, we obtain

$$0 \le \|x - tz\|^2 = \|x\|^2 - \|y\|^{-2}|\langle x, y\rangle|$$

with equality iff $x - tz = x - \alpha ty = 0$, from which the desired result is immediate. □

(5.20) Proposition. The function $x \to \|x\|$ is a norm on $\mathscr{X}$.

Proof. That $\|x\| = 0$ iff $x = 0$ and that $\|\lambda x\| = |\lambda|\|x\|$ are obvious from the definitions. As for the triangle inequality, the Schwarz inequality implies that

$$\|x + y\|^2 = \|x\|^2 + 2\operatorname{Re}\langle x, y\rangle + \|y\|^2 \le \|x\|^2 + 2\|x\|\|y\| + \|y\|^2$$

$$= (\|x\| + \|y\|)^2.$$ □

A pre-Hilbert space which is complete with respect to the norm $\|x\| = \sqrt{\langle x, x\rangle}$ is called a **Hilbert space**. (One can also consider real Hilbert spaces, with real scalar products. However, Hilbert spaces are usually assumed to be complex unless otherwise specified.)

Example. Let $(X, \mathscr{M}, \mu)$ be a measure space, and let $L^2(\mu)$ be the set of all measurable functions $f: X \to \mathbf{C}$ such that $\int |f|^2\, d\mu < \infty$ (where, as usual, two functions are identified if they are equal a.e.). From the inequality $ab \le \frac{1}{2}(a^2 + b^2)$, valid for all $a, b \ge 0$, we see that if $f, g \in L^2(\mu)$ then $|f\bar{g}| \le \frac{1}{2}(|f|^2 + |g|^2)$, so that $f\bar{g} \in L^1(\mu)$. It follows easily that the formula

$$\langle f, g\rangle = \int f\bar{g}\, d\mu$$

defines a scalar product on $L^2(\mu)$. In fact, $L^2(\mu)$ is a Hilbert space for any measure μ. We shall prove completeness in Theorem (6.6) below; for the present we shall take this result for granted.

For the remainder of this section, $\mathscr{X}$ will denote a Hilbert space.

(5.21) Proposition. If $x_n \to x$ and $y_n \to y$, then $\langle x_n, y_n\rangle \to \langle x, y\rangle$.

Proof. By the Schwarz inequality,

$$|\langle x_n, y_n\rangle - \langle x, y\rangle| = |\langle x_n - x, y_n\rangle - \langle x, y_n - y\rangle|$$

$$\le \|x_n - x\|\|y_n\| + \|x\|\|y_n - y\| \to 0.$$ □

(5.22) The Parallelogram Law. For all $x, y \in \mathscr{X}$, $\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2)$. ("The sum of the squares of the diagonals of a parallelogram is the sum of the squares of the sides.")

Proof. Expanding the scalar products defining $\|x + y\|$ and $\|x - y\|$, we find

$$\|x + y\|^2 + \|x - y\|^2 = \left(\|x\|^2 + 2\operatorname{Re}\langle x, y\rangle + \|y\|^2\right)$$

$$+\left(\|x\|^2 - 2\operatorname{Re}\langle x, y\rangle + \|y\|^2\right)$$

$$= 2\left(\|x\|^2 + \|y\|^2\right).$$ □

If $x, y \in \mathscr{X}$, we say that x is **orthogonal** to y and write $x \perp y$ if $\langle x, y\rangle = 0$. If $A \subset \mathscr{X}$, we define

$$A^\perp = \{x \in \mathscr{X}: \langle x, y\rangle = 0 \text{ for all } y \in A\}.$$

It is immediate from Proposition (5.21) and the linearity of the scalar product that $A^\perp$ is a closed subspace of $\mathscr{X}$.

(5.23) The Pythagorean Theorem. If $x_1, \ldots, x_n \in \mathscr{X}$ and $x_j \perp x_k$ for $j \neq k$, then $\|\sum_1^n x_j\|^2 = \sum_1^n \|x_j\|^2$.

Proof.

$$\left\|\sum x_j\right\|^2 = \left\langle \sum x_j, \sum x_j\right\rangle = \sum_{j,k} \langle x_j, x_k\rangle = \sum \|x_j\|^2.$$ □

(5.24) Theorem. If $\mathscr{M}$ is a closed subspace of $\mathscr{X}$, then $\mathscr{X} = \mathscr{M} \oplus \mathscr{M}^\perp$; that is, each $x \in \mathscr{X}$ can be expressed uniquely as $x = y + z$ where $y \in \mathscr{M}$ and $z \in \mathscr{M}^\perp$. Moreover, y and z are the unique elements of $\mathscr{M}$ and $\mathscr{M}^\perp$ whose distance to x is minimal.

Proof. Given $x \in \mathscr{X}$, let $\delta = \inf\{\|x - y\|: y \in \mathscr{M}\}$, and let $\{y_n\}$ be a sequence in $\mathscr{M}$ such that $\|x - y_n\| \to \delta$. By the parallelogram law,

$$2\big(\|y_n - x\|^2 + \|y_m - x\|^2\big) = \|y_n - y_m\|^2 + \|y_n + y_m - 2x\|^2,$$

or, since $\frac{1}{2}(y_n + y_m) \in \mathscr{M}$,

$$\|y_n - y_m\|^2 = 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\|\tfrac{1}{2}(y_n - y_m) - x\|^2$$

$$\leq 2\big(\|y_n - x\|^2 + \|y_m - x\|^2\big) - 4\delta^2.$$

As $m, n \to \infty$ this last quantity tends to zero, so $\{y_n\}$ is a Cauchy sequence. Let $y = \lim y_n$ and $z = x - y$. Then $y \in \mathscr{M}$ since $\mathscr{M}$ is closed, and $\|x - y\| = \delta$. We claim that $z \in \mathscr{M}^\perp$. Indeed, if $u \in \mathscr{M}$, after multiplying u by a nonzero scalar we may assume that $\langle z, u\rangle$ is real. Then the function

$$f(t) = \|z + tu\|^2 = \|z\|^2 + 2t\langle z, u\rangle + t^2\|u\|^2$$

is real for $t \in \mathbf{R}$, and it has a minimum (namely δ^2) at $t = 0$, because $z + tu = x - (y - tu)$ and $y - tu \in \mathscr{M}$. Thus $2\langle z, u\rangle = f'(0) = 0$, so $z \in \mathscr{M}^\perp$. Moreover, if z' is another element of $\mathscr{M}^\perp$, by the Pythagorean theorem (since $x - z \in \mathscr{M}$) we have

$$\|x - z'\|^2 = \|x - z\|^2 + \|z - z'\|^2 \geq \|x - z\|^2$$

with equality iff $z = z'$. The same reasoning also shows that y is the unique element of $\mathscr{M}$ closest to x. Finally, if $x = y' + z'$ with $y' \in \mathscr{M}$ and $z' \in \mathscr{M}^\perp$, then $y - y' = z' - z \in \mathscr{M} \cap \mathscr{M}^\perp$, so $y - y'$ and $z' - z$ are orthogonal to themselves and hence are zero. □

If $y \in \mathscr{X}$, the Schwarz inequality shows that the formula $f_y(x) = \langle x, y\rangle$ defines a bounded linear functional on $\mathscr{X}$ such that $\|f_y\| = \|y\|$. Thus, the map $y \to f_y$ is a conjugate-linear isometry of $\mathscr{X}$ into $\mathscr{X}^*$. It is a fundamental fact that this map is surjective:

(5.25) Theorem. If $f \in \mathscr{X}^*$, there is a unique $y \in \mathscr{X}$ such that $f(x) = \langle x, y\rangle$ for all $x \in \mathscr{X}$.

Proof. Uniqueness is obvious: if $\langle x, y\rangle = \langle x, y'\rangle$ for all x, by taking $x = y - y'$ we conclude that $\|y - y'\|^2 = 0$. If f is the zero functional, then obviously $y = 0$. Otherwise, let $\mathscr{M} = \{x \in \mathscr{X}: f(x) = 0\}$. Then $\mathscr{M}$ is a proper

closed subspace of $\mathscr{X}$, so by Theorem (5.24) there exists $z \in \mathscr{M}^\perp$ with $\|z\| = 1$. If $x \in \mathscr{X}$ and $u = f(x)z - f(z)x$, then $u \in \mathscr{M}$, so

$$0 = \langle u, z\rangle = f(x)\|z\|^2 - f(z)\langle x, z\rangle,$$

which means that $f(x) = f(z)\langle x, z\rangle = \langle x, y\rangle$ where $y = \overline{f(z)}\, z$. □

Thus, Hilbert spaces are reflexive in a very strong sense: not only is $\mathscr{X}$ naturally isomorphic to $\mathscr{X}^{**}$, it is naturally isomorphic (via a conjugate-linear map) to $\mathscr{X}^*$.

A subset $\{u_\alpha\}_{\alpha \in A}$ of $\mathscr{X}$ is called **orthonormal** if $\|u_\alpha\| = 1$ for all α and $u_\alpha \perp u_\beta$ whenever $\alpha \neq \beta$. If $\{x_n\}_1^\infty$ is a linearly independent sequence in $\mathscr{X}$, there is a standard procedure, called the **Gram–Schmidt process**, for converting $\{x_n\}$ into an orthonormal sequence $\{u_n\}$ such that the linear span of $\{u_n\}_1^N$ equals the linear span of $\{x_n\}_1^N$ for all N. To begin with, one sets $u_1 = x_1/\|x_1\|$. Having defined $u_1, \ldots, u_{N-1}$, one sets $v_N = x_N - \sum_1^{N-1}\langle x_N, u_n\rangle u_n$. v_N is nonzero because x_N is not in the linear span of $x_1, \ldots, x_{N-1}$ and hence of $u_1, \ldots, u_{N-1}$, so one can set $u_n = v_n/\|v_n\|$. It is easily verified that $\{u_n\}_1^\infty$ has the desired properties.

(5.26) Bessel's Inequality. If $\{u_\alpha\}_{\alpha \in A}$ is an orthonormal set, then for any $x \in \mathscr{X}$,

$$\sum_{\alpha \in A} |\langle x, u_\alpha\rangle|^2 \leq \|x\|^2.$$

In particular, $\{\alpha: \langle x, u_\alpha\rangle \neq 0\}$ is countable.

Proof. It suffices to show that $\sum_{\alpha \in F}|\langle x, u_\alpha\rangle|^2 \leq \|x\|^2$ for any finite $F \subset A$. But, by the Pythagorean theorem,

$$0 \leq \Big\|x - \sum_{\alpha \in F}\langle x, u_\alpha\rangle u_\alpha\Big\|^2$$

$$\leq \|x\|^2 - 2\,\mathrm{Re}\Big\langle x, \sum_{\alpha \in F}\langle x, u_\alpha\rangle u_\alpha\Big\rangle + \Big\|\sum_{\alpha \in F}\langle x, u_\alpha\rangle u_\alpha\Big\|^2$$

$$= \|x\|^2 - 2\sum_{\alpha \in F}|\langle x, u_\alpha\rangle|^2 + \sum_{\alpha \in F}|\langle x, u_\alpha\rangle|^2$$

$$= \|x\|^2 - \sum_{\alpha \in F}|\langle x, u_\alpha\rangle|^2.$$ □

(5.27) Theorem. If $\{u_\alpha\}_{\alpha\in A}$ is an orthonormal set, the following are equivalent:

(a) (Completeness) If $\langle x, u_\alpha\rangle = 0$ for all α, then $x = 0$.

(b) (Parseval's equation) $\|x\|^2 = \sum_{\alpha\in A}|\langle x, u_\alpha\rangle|^2$ for all $x \in \mathscr{X}$.

(c) For each $x \in \mathscr{X}$, $x = \sum_{\alpha\in A}\langle x, u_\alpha\rangle u_\alpha$, where the sum on the right has only countably many nonzero terms and converges in norm to x no matter how these terms are ordered.

Proof. (a) implies (c): If $x \in \mathscr{X}$, let $\alpha_1, \alpha_2, \ldots$ be any enumeration of those α's for which $\langle x, u_\alpha\rangle \neq 0$. By Bessel's inequality the series $\sum|\langle x, u_{\alpha_j}\rangle|^2$ converges, so by the Pythagorean theorem,

$$\left\|\sum_n^m \langle x, u_{\alpha_j}\rangle u_{\alpha_j}\right\|^2 = \sum_n^m |\langle x, u_{\alpha_j}\rangle|^2 \to 0 \quad \text{as } m, n \to \infty.$$

The series $\sum_1^\infty \langle x, u_{\alpha_j}\rangle u_{\alpha_j}$ therefore converges. If $y = x - \sum_1^\infty \langle x, u_{\alpha_j}\rangle u_{\alpha_j}$, then clearly $\langle y, u_\alpha\rangle = 0$ for all α, so by (a), $y = 0$.

(c) implies (b): With notation as above, as in the proof of Bessel's inequality we have

$$\|x\|^2 - \sum_1^n |\langle x, u_{\alpha_j}\rangle|^2 = \left\|x - \sum_1^n \langle x, u_{\alpha_j}\rangle u_{\alpha_j}\right\|^2 \to 0 \quad \text{as } n \to \infty.$$

Finally, that (b) implies (a) is obvious. □

An orthonormal set having the properties (a)–(c) in Theorem (5.27) is called an **orthonormal basis** for $\mathscr{X}$.

(5.28) Proposition. Every Hilbert space has an orthonormal basis.

Proof. A routine application of Zorn's lemma shows that the collection of orthonormal sets, ordered by inclusion, has a maximal element; and maximality is equivalent to property (a) in Theorem (5.27). □

(5.29) Proposition. A Hilbert space $\mathscr{X}$ is separable iff it has a countable orthonormal basis, in which case every orthonormal basis for $\mathscr{X}$ is countable.

Proof. If $\{x_n\}$ is a countable dense set in $\mathscr{X}$, by discarding recursively any x_n which is in the linear span of $x_1, \ldots, x_{n-1}$ we obtain a linearly independent sequence $\{y_n\}$ whose linear span is dense in $\mathscr{X}$. Application of the Gram–Schmidt process to $\{y_n\}$ yields an orthonormal sequence $\{x_n\}$ whose linear span is dense in $\mathscr{X}$ and which is therefore a basis. Conversely, if $\{u_n\}$ is a countable orthonormal basis, the finite linear combinations of the u_n's with coefficients in a countable dense subset of $\mathbf{C}$ form a countable dense set in $\mathscr{X}$. Moreover, if $\{v_\alpha\}_{\alpha\in A}$ is another orthonormal basis, for each n the set

$A_n = \{\alpha \in A: \langle u_n, v_\alpha\rangle \neq 0\}$ is countable. By completeness of $\{u_n\}$, $A = \bigcup_1^\infty A_n$, so A is countable. □

Most Hilbert spaces that arise in practice are separable. We discuss some examples in Exercises 60–62.

If A is an arbitrary set, let us denote by $l^2(A)$ the space $L^2(\mu)$ where μ is counting measure on $(A, \mathscr{P}(A))$. For each $\alpha \in A$ let $e_\alpha \in l^2(A)$ be defined by $e_\alpha(\beta) = 1$ if $\beta = \alpha$, $e_\alpha(\beta) = 0$ otherwise. Then for any $f \in l^2(A)$ we have $\langle f, e_\alpha\rangle = f(\alpha)$, and it follows easily that $\{e_\alpha\}_{\alpha\in A}$ is an orthonormal basis for $l^2(A)$. On the other hand, let $\mathscr{X}$ be an arbitrary Hilbert space with orthonormal basis $\{u_\alpha\}_{\alpha\in A}$. For each $x \in \mathscr{X}$, define $\hat{x}$: $A \to \mathbf{C}$ by $\hat{x}(\alpha) = \langle x, u_\alpha\rangle$. Thus $\hat{x} = \sum\hat{x}(\alpha)u_\alpha$—in particular, $\hat{u}_\alpha = e_\alpha$—and the Parseval equation shows that the map $x \to \hat{x}$ is an isometry from $\mathscr{X}$ into $l^2(A)$. In fact, this map is an isomorphism: if $f \in l^2(A)$, then $\sum|f(\alpha)|^2 < \infty$, so the Pythagorean theorem shows that the partial sums of the series $\sum f(\alpha)u_\alpha$ (of which only countably many terms are nonzero) are Cauchy; hence $x = \sum f(\alpha)u_\alpha$ exists in $\mathscr{X}$ and $\hat{x} = f$. Moreover, for $x, y \in \mathscr{X}$, by Proposition (5.21) we have

$$\langle x, y\rangle = \langle\sum\hat{x}(\alpha)u_\alpha, \sum\hat{y}(\alpha)u_\alpha\rangle = \sum\hat{x}(\alpha)\overline{\hat{y}(\alpha)} = \langle\hat{x}, \hat{y}\rangle.$$

If $\mathscr{X}$ and $\mathscr{Y}$ are Hilbert spaces, a linear isomorphism T: $\mathscr{X} \to \mathscr{Y}$ is called **unitary** if $\langle Tx, Ty\rangle = \langle x, y\rangle$ for all $x, y \in \mathscr{X}$. Summing up, we have proved:

(5.30) Proposition. Every Hilbert space is unitarily isomorphic to $l^2(A)$ for some set A.

EXERCISES

In these exercises, $\mathscr{X}$ denotes a Hilbert space.

55. (The polarization identity) For any $x, y \in \mathscr{X}$,

$$4\langle x, y\rangle = \|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2.$$

56. If $E \subset \mathscr{X}$, $(E^\perp)^\perp$ is the smallest closed subspace of $\mathscr{X}$ containing E.

57. Suppose $T \in L(\mathscr{X}, \mathscr{X})$.

(a) There is a unique $T^* \in L(\mathscr{X}, \mathscr{X})$ such that $\langle Tx, y\rangle = \langle x, T^*y\rangle$ for all $x, y \in \mathscr{X}$. (T^* is called the **adjoint** of T. Cf. Exercise 26.)

(b) $\|T^*\| = \|T\|$, $\|T^*T\| = \|T\|^2$, $(aS + bT)^* = \bar{a}S^* + \bar{b}T^*$, $(ST)^* = T^*S^*$, and $T^{**} = T$.

(c) Let $\mathscr{R}$ and $\mathscr{N}$ denote range and nullspace: then $\mathscr{R}(T)^\perp = \mathscr{N}(T^*)$ and $\mathscr{N}(T)^\perp = \overline{\mathscr{R}(T^*)}$.

(d) T is unitary iff T is invertible and $T^{-1} = T^*$.

58. Let $\mathscr{M}$ be a closed subspace of $\mathscr{X}$, and for $x \in \mathscr{X}$ let Px be the unique element of $\mathscr{M}$ such that $x - Px \in \mathscr{M}^\perp$ as in Theorem (5.24).

(a) $P \in L(\mathscr{X}, \mathscr{X})$, and in the notation of Exercise 57 we have $P^* = P$, $P^2 = P$, $\mathscr{R}(P) = \mathscr{M}$, and $\mathscr{N}(P) = \mathscr{M}^\perp$. P is called the **orthogonal projection** onto $\mathscr{M}$.

(b) Conversely, suppose $P \in L(\mathscr{X}, \mathscr{X})$ satisfies $P^2 = P$ and $P = P^*$. Then $\mathscr{R}(P)$ is closed and P is the orthogonal projection onto $\mathscr{R}(P)$.

(c) If $\{u_\alpha\}$ is an orthonormal basis for $\mathscr{M}$, then $Px = \Sigma\langle x, u_\alpha\rangle u_\alpha$.

59. Every closed convex set $K \subset \mathscr{X}$ has a unique element of minimal norm. [If $0 \in K$, the result is trivial; otherwise, adapt the proof of Theorem (5.24).]

60. Let $(X, \mathscr{M}, \mu)$ be a measure space. If $E \in \mathscr{M}$, we identify $L^2(E, \mu)$ with the subspace of $L^2(X, \mu)$ consisting of functions which vanish outside E. If $\{E_n\}$ is a disjoint sequence in $\mathscr{M}$ with $X = \bigcup_1^\infty E_n$, then $\{L^2(E_n, \mu)\}$ is a sequence of mutually orthogonal subspaces of $L^2(X, \mu)$, and every $f \in L^2(X, \mu)$ can be written uniquely as $f = \Sigma_1^\infty f_n$ (the series converging in norm) where $f_n \in L^2(E_n, \mu)$. If $L^2(E_n, \mu)$ is separable for all n, so is $L^2(X, \mu)$.

61. Let $(X, \mathscr{M}, \mu)$ and $(Y, \mathscr{N}, \nu)$ be σ-finite measure spaces such that $L^2(\mu)$ and $L^2(\nu)$ are separable. If $\{f_m\}$ and $\{g_n\}$ are orthonormal bases for $L^2(\mu)$ and $L^2(\nu)$, and $h_{mn}(x, y) = f_m(x)g_n(y)$, then $\{h_{mn}\}$ is an orthonormal basis for $L^2(\mu \times \nu)$.

62. In this exercise the measure defining the L^2 spaces is Lebesgue measure.

(a) $C([0, 1])$ is dense in $L^2([0, 1])$. [Adapt the proof of Theorem (2.26).]

(b) The set of polynomials is dense in $L^2([0, 1])$.

(c) $L^2([0, 1])$ is separable.

(d) $L^2(\mathbf{R})$ is separable. (Use Exercise 60.)

(e) $L^2(\mathbf{R}^n)$ is separable. (Use Exercise 61.)

63. Let $\mathscr{X}$ be an infinite-dimensional Hilbert space.

(a) Every orthonormal sequence in $\mathscr{X}$ converges weakly to zero.

(b) The unit sphere $S = \{x: \|x\| = 1\}$ is weakly dense in the unit ball $B = \{x: \|x\| \leq 1\}$. (In fact, every $x \in B$ is the weak limit of a sequence in S.)

64. Let $\mathscr{X}$ be a separable infinite-dimensional Hilbert space with orthonormal basis $\{u_n\}_1^\infty$.

(a) Define $T_k \in L(\mathscr{X}, \mathscr{X})$ by $T_k(\Sigma a_n u_n) = a_k u_k$. Then $T_k \to 0$ strongly but not in the norm topology.

(b) Define $S_k \in L(\mathscr{X}, \mathscr{X})$ by $S_k(\Sigma a_n u_n) = \Sigma a_n u_{n+k}$. Then $S_k \to 0$ weakly but not strongly.

65. $l^2(A)$ is unitarily isomorphic to $l^2(B)$ iff $\text{card}(A) = \text{card}(B)$.

66. Let $\mathscr{M}$ be a closed subspace of $L^2([0, 1], m)$ which is contained in $C([0, 1])$.

(a) There exists $A > 0$ such that $\|f\|_u \leq C\|f\|_{L^2}$ for all $f \in \mathscr{M}$. (Use the closed graph theorem.)

(b) For each $x \in [0, 1]$ there exists $g_x \in \mathscr{M}$ such that $f(x) = \langle f, g_x\rangle$ for all $f \in \mathscr{M}$, and $\|g_x\|_{L^2} \leq A$.

(c) The dimension of $\mathscr{M}$ is at most A^2. (If $\{f_j\}$ is an orthonormal sequence in $\mathscr{M}$, $\Sigma_1^n |f_j(x)|^2 \leq A^2$ for all $x \in [0, 1]$.)

67. (The mean ergodic theorem) Let U be a unitary operator on the Hilbert space $\mathscr{X}$, $\mathscr{M} = \{x: Ux = x\}$, P the orthogonal projection onto $\mathscr{M}$ (Exercise 58), and $S_n = n^{-1}\Sigma_0^{n-1} U^j$. Then $S_n \to P$ strongly. (If $x \in \mathscr{M}$, then $S_n x = x$; if $x = y - Uy$ for some y, then $S_n x \to 0$. By Exercise 57d, $\mathscr{M} = \{x: U^* x = x\}$. Apply Exercise 57c with $T = I - U$.)

6. NOTES AND REFERENCES

Functional analysis is a vast subject of which we have barely scratched the surface here. For the reader who wishes to learn more, Reed and Simon [78] and Rudin [89] are good places to start; one should also familiarize oneself with the treatise of Dunford and Schwartz [26].

Functional analysis has roots in a number of classical problems, particularly in the theory of differential and integral equations. The study of particular infinite-dimensional function spaces began in earnest around 1907 with work of F. Riesz, Fréchet, Schmidt, Helly, and others, and the notion of an abstract normed vector space appeared in papers by several authors about 1920. The research of the succeeding decade culminated in Banach's classic book [6], which marked the emergence of functional analysis as an established discipline. Detailed historical accounts of the subject can be found in Dieudonné [23] and in the notes in Dunford and Schwartz [26].

Section 1. The integral for vector-valued functions developed in Exercises 15–20 is called the **Bochner integral**. The hypothesis that $\mathscr{Y}$ is separable can be omitted, but the functions in $L^1_{\mathscr{Y}}$ must then be required to have separable range (after modification on a null set). A more detailed account can be found in Cohn [19].

Section 3. The open mapping and closed graph theorems are due to Banach [6]. The uniform boundedness principle, as we have stated it, is due to Banach and Steinhaus [7]; however, the second part of the theorem—that if $\mathscr{X}$ is complete and $\sup_{T \in \mathscr{A}} \|Tx\| < \infty$ for all $x \in \mathscr{X}$, then $\sup_{T \in \mathscr{A}} \|T\| < \infty$—had been proved previously without using the Baire category theorem by what Dieudonné [23] calls the "method of the gliding hump." This rather pretty

(and elementary) argument has been largely neglected in recent years, but a modern exposition of it can be found in Hennefeld [50].

It is simple to construct unbounded linear maps $T\colon \mathscr{X} \to \mathscr{Y}$ when $\mathscr{X}$ is incomplete (see Exercises 33 and 34), but virtually impossible to do so when $\mathscr{X}$ is complete without using the axiom of choice. The standard method is the following: start with an unbounded $T\colon \mathscr{X} \to \mathscr{Y}$ where $\mathscr{X}$ is incomplete, and let $\bar{\mathscr{X}}$ be the completion of $\mathscr{X}$. Pick a basis $\{u_\alpha\}_{\alpha \in A}$ for $\mathscr{X}$ (meaning that every $x \in \mathscr{X}$ is a *finite* linear combination of the u_α's), and extend it to a basis $\{u_\alpha\}_{\alpha \in B}$ $(B \supset A)$ for $\bar{\mathscr{X}}$. (This is where the axiom of choice comes in.) Let $\mathscr{M}$ be the linear span of $\{u_\alpha\}_{\alpha \in B \setminus A}$, so that each $x \in \bar{\mathscr{X}}$ can be written uniquely as $x = x_1 + x_2$ where $x_1 \in \mathscr{X}$, $x_2 \in \mathscr{M}$. Then T can be extended to $\bar{\mathscr{X}}$ by setting $T(x_1 + x_2) = Tx_1$.

Section 4. Alaoglu's theorem, which was first announced in [2] and proved in detail in [3], supersedes a number of earlier results dealing with special cases. It was discovered independently by Bourbaki [13].

Section 5. The space envisaged by Hilbert himself was $l^2(\mathbf{N})$; the notion of an abstract Hilbert space was introduced by von Neumann [73] in his work on the mathematics of quantum mechanics. Theorem (5.25) is originally due to F. Riesz [81] in the setting of L^2 spaces. It is one of several representation theorems for linear functionals on various spaces which bear his name, the others being Theorems (6.15), (7.2), and (7.17) below. To avoid confusion we reserve the name "Riesz representation theorem" for the latter two, which are closely related.

In the literature of quantum physics, scalar products are customarily denoted by $\langle x|y\rangle$ and are taken to be linear in the second variable and conjugate linear in the first.

6

L^p SPACES

L^p spaces are a class of Banach spaces of functions whose norms are defined in terms of integrals, and which generalize the L^1 spaces discussed in Chapter 2. They furnish interesting examples of the general theory of Chapter 5 and play a significant role in modern analysis.

1. BASIC THEORY OF L^p SPACES

In this chapter we shall be working on a fixed measure space $(X, \mathscr{M}, \mu)$. If f is a measurable function on X and $0 < p < \infty$, we define

$$\|f\|_p = \left(\int |f|^p \, d\mu\right)^{1/p}$$

(allowing the possibility that $\|f\|_p = \infty$), and we define

$$L^p(X, \mathscr{M}, \mu) = \{f\colon X \to \mathbf{C}\colon f \text{ is measurable and } \|f\|_p < \infty\}.$$

We abbreviate $L^p(X, \mathscr{M}, \mu)$ by $L^p(\mu)$, $L^p(X)$, or simply L^p when this will cause no confusion. As we have done with L^1, we consider two functions to define the same element of L^p when they are equal almost everywhere. If A is any nonempty set, we define $l^p(A)$ to be $L^p(\mu)$ where μ is counting measure on $(A, \mathscr{P}(A))$, and we denote $l^p(\mathbf{N})$ simply by l^p.

L^p is a vector space, for if $f, g \in L^p$, then

$$|f + g|^p \le [2\max(|f|, |g|)]^p \le 2^p(|f|^p + |g|^p).$$

Our notation suggests that $\| \ \|_p$ is a norm on L^p. Indeed, it is obvious that $\|f\|_p = 0$ iff $f = 0$ a.e. and $\|cf\|_p = |c|\|f\|_p$, so the only question is the triangle inequality. It turns out that the latter is valid precisely when $p \ge 1$, so our attention will be focused almost exclusively on this case.

Before proceeding further, let us see why the triangle inequality fails for $p < 1$. Suppose $a > 0$, $b > 0$, and $0 < p < 1$. For $t > 0$ we have $t^{p-1} > (a + t)^{p-1}$, and by integrating from 0 to b we obtain $a^p + b^p > (a + b)^p$. Then if E and F are disjoint sets of positive finite measure in X and we set $a = \mu(E)^{1/p}$, $b = \mu(F)^{1/p}$, we see that

$$\|\chi_E + \chi_F\|_p = (a^p + b^p)^{1/p} > a + b = \|\chi_E\|_p + \|\chi_F\|_p.$$

The cornerstone of the theory of L^p spaces is Hölder's inequality, which we now derive.

(6.1) Lemma. If $a \geq 0$, $b \geq 0$, and $0 < \lambda < 1$, then

$$a^\lambda b^{1-\lambda} \leq \lambda a + (1 - \lambda)b,$$

with equality iff $a = b$.

Proof. The result is obvious if $b = 0$; otherwise, setting $t = a/b$, we need to show that $t^\lambda \leq \lambda t + (1 - \lambda)$ with equality iff $t = 1$. But by elementary calculus, $t^\lambda - \lambda t$ is strictly increasing for $t < 1$ and strictly decreasing for $t > 1$, so its maximum value, namely $1 - \lambda$, occurs at $t = 1$. □

(6.2) Hölder's Inequality. Suppose $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$ [in other words, $q = p/(p - 1)$]. If f and g are measurable functions on X,

$$\|fg\|_1 \leq \|f\|_p \|g\|_q. \tag{6.3}$$

In particular, if $f \in L^p$ and $g \in L^q$, then $fg \in L^1$, and in this case equality holds in (6.3) iff $\alpha|f|^p = \beta|g|^q$ a.e. for some constants α, β with $\alpha\beta \neq 0$.

Proof. The result is trivial if $\|f\|_p = 0$ or $\|g\|_q = 0$ (since then $f = 0$ or $g = 0$ a.e.), or if $\|f\|_p = \infty$ or $\|g\|_q = \infty$. Otherwise, we apply Lemma (6.1) with

$$a = \left|\frac{f(x)}{\|f\|_p}\right|^p, \quad b = \left|\frac{g(x)}{\|g\|_q}\right|^q, \quad \text{and } \lambda = \frac{1}{p}$$

to obtain

$$\frac{|f(x)g(x)|}{\|f\|_p \|g\|_q} \leq \frac{|f(x)|^p}{p\int |f|^p \, d\mu} + \frac{|g(x)|^q}{q\int |g|^q \, d\mu}. \tag{6.4}$$

Integration of both sides yields

$$\frac{\|fg\|_1}{\|f\|_p \|g\|_q} \leq \frac{1}{p} + \frac{1}{q} = 1.$$

Equality holds here iff it holds a.e. in (6.4), and by Lemma (6.1) this happens precisely when $\|g\|_q^q |f|^p = \|f\|_p^p |g|^q$ a.e. □

The equation $p^{-1} + q^{-1} = 1$ occurring in Hölder's inequality turns up frequently in L^p theory. If $1 < p < \infty$, the number $q = p/(p - 1)$ such that $p^{-1} + q^{-1} = 1$ is called the **conjugate exponent** to p.

(6.5) Minkowski's Inequality. If $1 \leq p < \infty$ and $f, g \in L^p$, then

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof. The result is obvious if $p = 1$ or if $f + g = 0$ a.e. Otherwise, we write

$$|f + g|^p \leq (|f| + |g|)|f + g|^{p-1}$$

and apply Hölder's inequality, observing that $(p - 1)q = p$ when q is the conjugate exponent to p:

$$\int |f + g|^p \, d\mu \leq \|f\|_p \| |f + g|^{p-1} \|_q + \|g\|_p \| |f + g|^{p-1} \|_q$$

$$= (\|f\|_p + \|g\|_p) \left(\int |f + g|^p \, d\mu \right)^{1/q}.$$

Therefore,

$$\|f + g\|_p = \left(\int |f + g|^p \, d\mu \right)^{1 - (1/q)} \leq \|f\|_p + \|g\|_p.$$

□

This result shows that, for $p \geq 1$, L^p is a normed vector space with norm $\| \|_p$. More is true:

(6.6) Theorem. For $1 \leq p < \infty$, L^p is a Banach space.

Proof. We use Theorem (5.1). Suppose $\{f_k\} \subset L^p$ and $\sum_1^\infty \|f_k\|_p = B < \infty$. Let $G_n = \sum_1^n |f_k|$ and $G = \sum_1^\infty |f_k|$. Then $\|G_n\|_p \leq \sum_1^n \|f_k\|_p \leq B$ for all n, so by the monotone convergence theorem, $\int G^p = \lim \int G_n^p \leq B^p$. Hence $G \in L^p$, and in particular $G(x) < \infty$ a.e., which implies that the series $\sum_1^\infty f_k$ converges a.e. Denoting its sum by F, we have $|F| \leq G$ and hence $F \in L^p$; moreover, $|F - \sum_1^n f_k|^p \leq (2G)^p \in L^1$, so by the dominated convergence theorem,

$$\left\| F - \sum_1^n f_k \right\|_p^p = \int \left| F - \sum_1^n f_k \right|^p \to 0.$$

Thus the series $\sum_1^\infty f_k$ converges in the L^p norm. □

(6.7) Proposition. For $1 \le p < \infty$, the set of simple functions $f = \sum_1^n a_j \chi_{E_j}$, where $\mu(E_j) < \infty$ for all j, is dense in L^p.

Proof. Clearly such functions are in L^p. If $f \in L^p$, choose a sequence $\{f_n\}$ of simple functions such that $f_n \to f$ a.e. and $|f_n| \le |f|$, according to Theorem (2.10). Then $f_n \in L^p$ and $|f_n - f|^p \le 2^p |f|^p \in L^1$, so by the dominated convergence theorem, $\|f_n - f\|_p \to 0$. Moreover, if $f_n = \sum a_j \chi_{E_j}$ where the E_j's are disjoint, and $a_j \ne 0$, we must have $\mu(E_j) < \infty$ since $\sum |a_j|^p \mu(E_j) = \int |f_n|^p < \infty$. □

To complete the picture of L^p spaces, we introduce a space corresponding to the limiting value $p = \infty$. If f is a measurable function on X, we define

$$\|f\|_\infty = \inf\{a \ge 0: \mu(\{x: |f(x)| > a\}) = 0\},$$

with the convention that $\inf \varnothing = \infty$. We observe that the infimum is actually attained, for

$$\{x: |f(x)| > a\} = \bigcup_1^\infty \{x: |f(x)| > a + n^{-1}\},$$

and if the sets on the right are null, so is the one on the left. $\|f\|_\infty$ is sometimes called the **essential supremum** of f and written

$$\|f\|_\infty = \operatorname{ess\,sup}_{x \in X} |f(x)|.$$

We now define

$$L^\infty = L^\infty(X, \mathcal{M}, \mu) = \{f: X \to \mathbf{C}: f \text{ is measurable and } \|f\|_\infty < \infty\},$$

with the usual convention that two functions are equal a.e. define the same element of L^∞. Thus $f \in L^\infty$ iff there is a bounded measurable function g such that $f = g$ a.e., for instance, $g = f\chi_E$ where $E = \{x: |f(x)| \le \|f\|_\infty\}$. [We remark that for fixed X and $\mathcal{M}$, $L^\infty(X, \mathcal{M}, \mu)$ depends on μ only insofar as μ determines which sets have measure zero. If μ is not semifinite, it is appropriate for some purposes to adopt a slightly different definition of L^∞; this point is explored in Exercises 23–26.]

The results we have proved above for $1 \le p < \infty$ extend easily to the case $p = \infty$. We sum up:

(6.8) Theorem.

(a) If f and g are measurable functions on X, then $\|fg\|_1 \le \|f\|_1 \|g\|_\infty$. If $f \in L^1$ and $g \in L^\infty$, $\|fg\|_1 = \|f\|_1 \|g\|_\infty$ iff $|g(x)| = \|g\|_\infty$ a.e. on the set where $f(x) \ne 0$.

(b) $\| \ \|_\infty$ is a norm on L^∞.

(c) $\|f_n - f\|_\infty \to 0$ iff there exists $E \in \mathcal{M}$ such that $\mu(E^c) = 0$ and $f_n \to f$ uniformly on E.

(d) L^∞ is a Banach space.

(e) The bounded simple functions are dense in L^∞.

The proof is left to the reader (Exercise 2).

In view of Theorem (6.8a) and the formal equality $1^{-1} + \infty^{-1} = 1$, it is natural to regard 1 and ∞ as conjugate exponents, and we do so henceforth.

Theorem (6.8c) shows that $\| \ \|_\infty$ is closely related to, but usually not identical with, the uniform norm $\| \ \|_u$. However, if we are dealing with Lebesgue measure, or more generally any Borel measure which assigns positive values to all open sets, then $\|f\|_\infty = \|f\|_u$ whenever f is continuous, since $\{x: |f(x)| > a\}$ is open. In this case we may use the notations $\|f\|_\infty$ and $\|f\|_u$ interchangeably, and we may regard the space of bounded continuous functions as a (closed!) subspace of L^∞.

In general we have $L^p \not\subset L^q$ for all $p \ne q$; to see what is at issue it is instructive to consider the following simple examples on $(0, \infty)$ with Lebesgue measure. Let $f_\alpha(x) = x^{-\alpha}$, where $\alpha > 0$. Elementary calculus shows that $f_\alpha \chi_{(0,1)} \in L^p$ iff $p < \alpha^{-1}$, and $f_\alpha \chi_{(1,\infty)} \in L^p$ iff $p > \alpha^{-1}$. Thus we see two reasons why a function f may fail to be in L^p: either $|f|^p$ blows up too rapidly near some point, or it fails to decay sufficiently rapidly at infinity. In the first situation the behavior of $|f|^p$ becomes worse as p increases, whereas in the second it becomes better. In other words, if $p < q$, functions in L^p can be locally more singular than functions in L^q, while functions in L^q can be globally more spread out than functions in L^p. These somewhat imprecisely expressed ideas are actually a rather accurate guide to the general situation, concerning which we now give four precise results. The last two show that inclusions $L^p \subset L^q$ can be obtained under conditions on the measure space which disallow one of the types of bad behavior described above; for a more general result, see Exercise 5.

(6.9) Proposition. If $0 < p < q < r \le \infty$, then $L^q \subset L^p + L^r$; that is, each $f \in L^q$ is the sum of a function in L^p and a function in L^r.

Proof. If $f \in L^q$, let $E = \{x: |f(x)| > 1\}$ and set $g = f\chi_E$, $h = f\chi_{E^c}$. Then $|g|^p = |f|^p \chi_E \le |f|^q \chi_E$, so $g \in L^p$, and $|h|^r = |f|^r \chi_{E^c} \le |f|^q \chi_{E^c}$, so $h \in L^r$. (If $r = \infty$, obviously $\|h\|_\infty = 1$.) □

(6.10) Proposition. If $0 < p < q < r \le \infty$, then $L^p \cap L^r \subset L^q$ and

$$\|f\|_q \le \|f\|_p^\lambda \|f\|_r^{1-\lambda}, \quad \text{where } \frac{1}{q} = \frac{\lambda}{p} + \frac{1-\lambda}{r}.$$

Proof. If $r = \infty$, we have $\int |f|^q \le \|f\|_\infty^{q-p} \int |f|^p$, so

$$\|f\|_q \le \|f\|_p^{p/q} \|f\|_\infty^{1-(p/q)} = \|f\|_p^\lambda \|f\|_\infty^{1-\lambda}.$$

If $r < \infty$, we use Hölder's inequality, taking the pair of conjugate exponents to

be $p/\lambda q$ and $r/(1-\lambda)q$:

$$\int |f|^q = \int |f|^{\lambda q}|f|^{(1-\lambda)q} \le \||f|^{\lambda q}\|_{p/\lambda q}\||f|^{(1-\lambda)q}\|_{r/(1-\lambda)q}$$

$$= \left(\int |f|^p\right)^{\lambda q/p}\left(\int |f|^r\right)^{(1-\lambda)q/r}$$

$$= \|f\|_p^{\lambda q}\|f\|_r^{(1-\lambda)q}.$$

Taking q-th roots, we are done. □

(6.11) Proposition. If A is any set and $0 < p < q \le \infty$, then $l^p(A) \subset l^q(A)$, and $\|f\|_q \le \|f\|_p$.

Proof. Obviously $\|f\|_\infty^p = \sup_\alpha |f(\alpha)|^p \le \sum_\alpha |f(\alpha)|^p$, so that $\|f\|_\infty \le \|f\|_p$. The case $q < \infty$ then follows from Proposition (6.10): if $\lambda = p/q$,

$$\|f\|_q \le \|f\|_p^\lambda \|f\|_\infty^{1-\lambda} \le \|f\|_p. \qquad \square$$

(6.12) Proposition. If $\mu(X) < \infty$ and $0 < p < q \le \infty$, then $L^p(\mu) \supset L^q(\mu)$ and $\|f\|_p \le \|f\|_q \mu(X)^{(1/p)-(1/q)}$.

Proof. If $q = \infty$, this is obvious:

$$\|f\|_p^p = \int |f|^p \le \|f\|_\infty^p \int 1 = \|f\|_\infty^p \mu(X).$$

If $q < \infty$, we use Hölder's inequality with the conjugate exponents q/p and $q/(q-p)$:

$$\|f\|_p^p = \int |f|^p \cdot 1 \le \||f|^p\|_{q/p}\|1\|_{q/(q-p)} = \|f\|_q^p \mu(X)^{(q-p)/q}. \qquad \square$$

We conclude this section with a few remarks about the significance of the L^p spaces. The three most obviously important ones are L^1, L^2, and L^∞. With L^1 we are already familiar; L^2 is special because it is a Hilbert space, and the topology of L^∞ is closely related to the topology of uniform convergence. Unfortunately, L^1 and L^∞ are pathological in many respects, and it is often more fruitful to deal with the intermediate L^p spaces. One manifestation of this is the duality theory in Section 2; another is the fact that many operators of interest in Fourier analysis and differential equations are bounded on L^p for $1 < p < \infty$ but not on L^1 or L^∞. (Some examples are mentioned in Section 8.9.)

EXERCISES

1. When does equality hold in Minkowski's inequality? (The answer is different for $p = 1$ and for $1 < p < \infty$. What about $p = \infty$?)
2. Prove Theorem (6.8).
3. If $1 \le p < r \le \infty$, $L^p \cap L^r$ is a Banach space with norm $\|f\| = \|f\|_p + \|f\|_r$, and if $p < q < r$, the inclusion map $L^p \cap L^r \to L^q$ is continuous.
4. If $1 \le p < r \le \infty$, $L^p + L^r$ is a Banach space with norm $\|f\| = \inf\{\|g\|_p + \|h\|_r : f = g + h\}$, and if $p < q < r$, the inclusion map $L^q \to L^p + L^r$ is continuous.
5. Suppose $0 < p < q \le \infty$. Then $L^p \not\subset L^q$ iff X contains sets of arbitrarily small positive measure, and $L^q \not\subset L^p$ iff X contains sets of arbitrarily large finite measure. [For the "if" implication: in the first case there is a disjoint sequence $\{E_n\}$ with $0 < \mu(E_n) \le 2^{-n}$, and in the second case there is a disjoint sequence $\{E_n\}$ with $1 \le \mu(E_n) < \infty$. Consider $f = \sum a_n \chi_{E_n}$ for suitable constants a_n.]
6. Suppose $0 < p_0 < p_1 \le \infty$. Find examples of functions f on $(0, \infty)$ (with Lebesgue measure) such that $f \in L^p$ iff (a) $p_0 < p < p_1$, (b) $p_0 \le p \le p_1$, (c) $p = p_0$. [Consider functions of the form $f(x) = x^{-\alpha}|\log x|^\beta$.]
7. If $f \in L^p \cap L^\infty$ for some $p < \infty$, so that $f \in L^q$ for $q > p$, then $\|f\|_\infty = \lim_{q\to\infty}\|f\|_q$.
8. Suppose $\mu(X) = 1$ and $f \in L^p$ for some $p > 0$, so that $f \in L^q$ for $0 < q < p$.
   (a) $\log\|f\|_q \ge \int \log|f|$. [Use Exercise 3.42d, with $F(t) = e^t$.]
   (b) $(\int |f|^q - 1)/q \ge \log\|f\|_q$, and $(\int |f|^q - 1)/q \to \int \log|f|$ as $q \to 0$.
   (c) $\lim_{q\to 0}\|f\|_q = \exp(\int \log|f|)$.
9. If $\|f_n - f\|_p \to 0$ where $p < \infty$, then $f_n \to f$ in measure, and hence some subsequence converges to f a.e. On the other hand, if $f_n \to f$ in measure and $|f_n| \le g \in L^p$ for all n where $p < \infty$, then $\|f_n - f\|_p \to 0$.
10. If $f_n, f \in L^p$ $(p < \infty)$ and $f_n \to f$ a.e., then $\|f_n - f\|_p \to 0$ iff $\|f_n\|_p \to \|f\|_p$. (Use Exercise 2.20.)
11. If f is a measurable function on x, define the **essential range** of f, R_f, to be the set of all $z \in \mathbf{C}$ such that $\{x: |f(x) - z| < \varepsilon\}$ has positive measure for all $\varepsilon > 0$. If $f \in L^\infty$, then R_f is compact and $\|f\|_\infty = \sup\{|z|: z \in R_f\}$.
12. If $p \ne 2$, the L^p norm does not arise from a scalar product on L^p. (Show that the parallelogram law fails.)
13. $L^p(\mathbf{R}, m)$ is separable for $1 \le p < \infty$. (Cf. Exercise 5.62.) However, $L^\infty(\mathbf{R}, m)$ is not separable. (There is an uncountable set $\mathscr{F} \subset L^\infty$ such that $\|f - g\|_\infty \ge 1$ for all $f, g \in \mathscr{F}$ with $f \ne g$.)
14. If $g \in L^\infty$, the operator T defined by $Tf = fg$ is bounded on L^p for $1 \le p \le \infty$. Its operator norm is at most $\|g\|_\infty$, with equality if μ is semifinite.

15. (The Vitali Convergence Theorem) Suppose $\{f_n\}_1^\infty \subset L^p$ where $1 \le p < \infty$. In order for $\{f_n\}$ to be Cauchy in the L^p norm it is necessary and sufficient for the following three conditions to hold: (i) $\{f_n\}$ is Cauchy in measure, (ii) the sequence $\{|f_n|^p\}$ is uniformly integrable (cf. Exercise 3.11), and (iii) for every $\varepsilon > 0$ there exists $E \subset X$ such that $\mu(E) < \infty$ and $\int_{E^c} |f_n|^p < \varepsilon$ for all n. [To prove the sufficiency, given $\varepsilon > 0$ let E be as in (iii), and let $A_{mn} = \{x \in E: |f_m(x) - f_n(x)| \ge \varepsilon\}$. Then the integrals of $|f_m - f_n|^p$ over $E \setminus A_{mn}$, A_{mn}, and E^c are small when m and n are large—for three different reasons.]

16. If $0 < p < 1$, the formula $\rho(f, g) = \int |f - g|^p$ defines a metric on L^p which makes L^p into a complete topological vector space. [The proof of Theorem (6.6) still works for $p < 1$ if $\|f\|_p$ is replaced by $\int |f|^p$, as it uses only the triangle inequality and not the homogeneity of the norm.]

2. THE DUAL OF L^p

Suppose p and q are conjugate exponents. Hölder's inequality shows that each $g \in L^q$ defines a bounded linear functional φ_g on L^p by

$$\varphi_g(f) = \int fg,$$

and the operator norm of φ_g is at most $\|g\|_q$. [If $p = 2$ and we are thinking of L^2 as a Hilbert space, it is more appropriate to define $\varphi_g(f) = \int f\bar{g}$. The same convention can be used for $p \ne 2$ without changing the results below in any essential way.] In fact, the map $g \to \varphi_g$ is almost always an isometry from L^q into $(L^p)^*$.

(6.13) Proposition. If p and q are conjugate exponents, $1 \le q < \infty$, and $g \in L^q$, then

$$\|g\|_q = \|\varphi_g\| = \sup\left\{\left|\int fg\right|: \|f\|_p = 1\right\}.$$

If μ is semifinite, this result holds also for $q = \infty$.

Proof. We have seen that $\|\varphi_g\| \le \|g\|_q$, and equality is trivial if $\|g\|_q = 0$. If $g \ne 0$ and $q < \infty$, let

$$f = \frac{|g|^{q-1}\overline{\operatorname{sgn} g}}{\|g\|_q^{q-1}}.$$

Then

$$\|f\|_p^p = \frac{\int |g|^{(q-1)p}}{\|g\|_q^{(q-1)p}} = \frac{\int |g|^q}{\int |g|^q} = 1,$$

and

$$\|\varphi_g\| \ge \int fg = \frac{\int |g|^q}{\|g\|_q^{q-1}} = \|g\|_q.$$

If $q = \infty$ and μ is semifinite, for each $\varepsilon > 0$ we can choose $A \subset \{x: |g(x)| > \|g\|_\infty - \varepsilon\}$ such that $0 < \mu(A) < \infty$. Let $f = \mu(A)^{-1}\chi_A \overline{\operatorname{sgn} g}$: then $\|f\|_1 = 1$ and

$$\|\varphi_g\| \ge \int fg = \mu(A)^{-1}\int_A |g| \ge \|g\|_\infty - \varepsilon. \qquad \square$$

Conversely, if μ is semifinite and $f \to \int fg$ is a bounded linear functional on L^p, then $f \in L^q$. In fact, we have the following slightly stronger result:

(6.14) Converse of Hölder's Inequality. Suppose μ is semifinite and p, q are conjugate exponents. If g is a measurable function on X such that $fg \in L^1$ for all f in the space S of simple functions which vanish outside a set of finite measure, and the quantity

$$M_q(g) = \sup\left\{\left|\int fg\right|: f \in S \text{ and } \|f\|_p = 1\right\}$$

is finite, then $g \in L^q$ and $M_q(g) = \|g\|_q$.

Proof. In the proof for $q < \infty$ we shall assume that μ is σ-finite, leaving the general case as Exercise 17. Let $\{E_n\}$ be an increasing sequence of sets of positive finite measure such that $X = \bigcup_1^\infty E_n$. Let $\{\varphi_n\}$ be a sequence of simple functions such that $\varphi_n \to g$ pointwise and $|\varphi_n| \le |g|$, and let $g_n = \varphi_n \chi_{E_n}$. Then $g_n \to g$ pointwise, $|g_n| \le |g|$, and $g_n \in S$. Let $f_n = \|g_n\|_q^{1-q}|g_n|^{q-1}\overline{\operatorname{sgn} g}$. Then as in the proof of Proposition (6.13) we have $\|f_n\|_p = 1$, and by Fatou's lemma,

$$\|g\|_q \le \liminf \|g_n\|_q = \liminf \int |f_n g_n|$$

$$\le \liminf \int |f_n g| = \liminf \int f_n g \le M_q(g).$$

If $q = \infty$ and $\varepsilon > 0$, let $A = \{x: |g(x)| \ge M_\infty(g) + \varepsilon\}$. If $\mu(A)$ were positive, we could choose $B \subset A$ with $0 < \mu(B) < \infty$; setting $f = \mu(B)^{-1}\overline{\operatorname{sgn} g}\chi_B$ we would then have $\|f\|_1 = 1$ and $\int fg \ge M_\infty(g) + \varepsilon$, which is absurd. Hence $\|g\|_\infty \le M_\infty(g)$.

Finally, Hölder's inequality gives $M_q(g) \le \|g\|_q$, so the proof is complete. $\square$

The last and deepest part of the description of $(L^p)^*$ is the fact that the map $g \to \varphi_g$ is, in almost all cases, a surjection.

(6.15) Theorem. Suppose p and q are conjugate exponents. If μ is σ-finite and $1 \le p < \infty$, for each $\varphi \in (L^p)^*$ there exists $g \in L^q$ such that $\varphi(f) = \int fg$ for all $f \in L^p$, and hence $(L^p)^*$ is isometrically isomorphic to L^q. If $1 < p < \infty$, this conclusion holds with no restriction on μ.

Proof. First let us suppose that μ is finite, so that all simple functions are in L^p. If $\varphi \in (L^p)^*$ and E is a measurable set, let $\nu(E) = \varphi(\chi_E)$. For any disjoint sequence $\{E_j\}$, if $E = \bigcup_1^\infty E_j$, we have $\chi_E = \sum_1^\infty \chi_{E_j}$ where the series converges in the L^p norm:

$$\left\| \chi_E - \sum_1^n \chi_{E_j} \right\|_p = \left\| \sum_{n+1}^\infty \chi_{E_j} \right\|_p = \mu\left(\bigcup_{n+1}^\infty E_j \right)^{1/p} \to 0 \text{ as } n \to \infty.$$

(It is at this point that we need the assumption that $p < \infty$.) Hence, since φ is linear and continuous,

$$\nu(E) = \sum_1^\infty \varphi(\chi_{E_j}) = \sum_1^\infty \nu(E_j),$$

so that ν is a complex measure. Also, if $\mu(E) = 0$, then $\chi_E = 0$ as an element of L^p, so $\nu(E) = 0$; that is, $\nu \ll \mu$. By the Radon–Nikodym theorem there exists $g \in L^1(\mu)$ such that $\varphi(\chi_E) = \nu(E) = \int_E g\,d\mu$ for all E, and hence $\varphi(f) = \int fg\,d\mu$ for all simple functions f. Moreover, $|\int fg| \le \|\varphi\| \|f\|_p$, so $g \in L^q$ by Theorem (6.14). Once we know this, it follows from Proposition (6.7) that $\varphi(f) = \int fg$ for all $f \in L^p$.

Now suppose that μ is σ-finite. Let $\{E_n\}$ be an increasing sequence of sets such that $0 < \mu(E_n) < \infty$ and $X = \bigcup_1^\infty E_n$, and let us agree to identify $L^p(E_n)$ and $L^q(E_n)$ with the subspaces of $L^p(X)$ and $L^q(X)$ consisting of functions which vanish outside E_n. The preceding argument shows that for each n there exists $g \in L^q(E_n)$ such that $\varphi(f) = \int fg_n$ for all $f \in L^p(E_n)$, and $\|g_n\|_q = \|(\varphi | L^p(E_n))\| \le \|\varphi\|$. g_n is unique modulo alterations on nullsets, so $g_n = g_m$ a.e. on E_n for $n < m$ and we can define g a.e. on X by setting $g = g_n$ on E_n. By the monotone convergence theorem, $\|g\|_q = \lim \|g_n\|_q \le \|\varphi\|$, so $g \in L^q$. Moreover, if $f \in L^p$, then $f\chi_{E_n} \to f$ in the L^p norm by the dominated convergence theorem, so $\varphi(f) = \lim \varphi(f\chi_{E_n}) = \lim \int_{E_n} fg_n = \int fg$.

Finally, suppose that μ is arbitrary and $p > 1$, so that $q < \infty$. As above, for each σ-finite set $E \subset X$ there is an essentially unique $g_E \in L^q(E)$ such that $\varphi(f) = \int fg_E$ for all $f \in L^p(E)$, and $\|g_E\|_q \le \|\varphi\|$. If F is σ-finite and $F \supset E$, then $g_F = g_E$ a.e. on E, so $\|g_F\|_q \ge \|g_E\|_q$. Let M be the supremum of $\|g_E\|_q$ as E ranges over all σ-finite sets, choose a sequence $\{E_n\}$ so that $\|g_{E_n}\|_q \to M$, and set $F = \bigcup_1^\infty E_n$. Then F is σ-finite and $\|g_F\|_q \ge \|g_{E_n}\|_q$ for all n, hence $\|g_F\|_q = M$. Now, if A is a σ-finite set containing F, we have

$$\int |g_F|^q + \int |g_{A \setminus F}|^q = \int |g_A|^q \le M^q = \int |g_F|^q,$$

and thus $g_A = 0$ a.e. on $A \setminus F$. (Here we use the fact that $q < \infty$.) But if $f \in L^p$, then $A = \{x: f(x) \ne 0\} \cup F$ is σ-finite, so $\varphi(f) = \int fg_A = \int fg_F$. Thus we may take $g = g_F$, and the proof is complete. □

(6.16) Corollary. If $1 < p < \infty$, L^p is reflexive.

We conclude with some remarks on the exceptional cases $p = 1$ and $p = \infty$. For any measure μ, the correspondence $g \to \varphi_g$ maps L^∞ into $(L^1)^*$, but in general it is neither injective nor surjective. Injectivity fails when μ is not semifinite. Indeed, if $E \subset X$ is a set of infinite measure which contains no subsets of positive finite measure, and $f \in L^1$, then $\{x: f(x) \ne 0\}$ is σ-finite and hence intersects E in a null set. It follows that $\varphi_{\chi_E} = 0$ although $\chi_E \ne 0$ in L^∞. This problem, however, can be remedied by redefining L^∞: see Exercises 23–25. The failure of surjectivity is more subtle and is best illustrated by an example.

Let X be an uncountable set, μ = counting measure on $(X, \mathcal{P}(X))$, $\mathcal{M}$ = the σ-algebra of countable or co-countable sets, and μ_0 the restriction of μ to $\mathcal{M}$. Every $f \in L^1(\mu)$ vanishes outside a countable set, so it follows that $L^1(\mu) = L^1(\mu_0)$. On the other hand, $L^\infty(\mu)$ consists of all bounded functions on X, whereas $L^\infty(\mu_0)$ consists of those bounded functions which are constant except on a countable set. With this in mind it is easy to see that the dual of $L^1(\mu_0)$ is $L^\infty(\mu)$ and not the smaller space $L^\infty(\mu_0)$. (See also Exercise 26.)

As for the case $p = \infty$: the map $g \to \varphi_g$ is always an injection of L^1 into $(L^\infty)^*$, but it is almost *never* a surjection. We shall say more about this in Section 6; for the present, we give a specific example. (Another example can be found in Exercise 19.)

Let $X = [0, 1]$, μ = Lebesgue measure. The map $f \to f(0)$ is a bounded linear functional on $C(X)$, which we regard as a subspace of L^∞. By the Hahn–Banach theorem there exists $\varphi \in (L^\infty)^*$ such that $\varphi(f) = f(0)$ for all $f \in C(X)$. To see that φ cannot be given by integration against an L^1 function, consider the functions $f_n \in C(X)$ defined by $f_n(x) = 1 - nx$ for $0 \le x \le n^{-1}$ and $f_n(x) = 0$ for $n^{-1} \le x \le 1$. By the dominated convergence theorem, $\int f_n g \to 0$ for all $g \in L^1$, whereas $\varphi(f_n) = 1$ for all n.

EXERCISES

17. Complete the proof of Theorem (6.14) for the non-σ-finite case. [If g satisfies the hypotheses with $q < \infty$, then $\mu(\{x: |g(x)| > n^{-1}\})$ must be finite for all n and hence $\{x: g(x) \ne 0\}$ is σ-finite.]

18. (von Neumann) The self-duality of L^2 follows from Hilbert space theory [Theorem (5.25)], and this fact can be used to prove the Lebesgue–Radon–Nikodym theorem. Suppose that μ, ν are positive finite

measures on $(X, \mathscr{M})$ (the σ-finite case follows as in Chapter 3), and let $\lambda = \mu + \nu$.

(a) The map $f \to \int f d\nu$ is a bounded linear functional on $L^2(\lambda)$, so $\int f d\nu = \int fg\, d\lambda$ for some $g \in L^2(\lambda)$. Equivalently, $\int f(1-g)\, d\nu = \int fg\, d\mu$ for $f \in L^2(\lambda)$.

(b) $0 \le g \le 1$ λ-a.e., so we may assume $0 \le g \le 1$ everywhere.

(c) Let $A = \{x: g(x) < 1\}$, $B = \{x: g(x) = 1\}$, and set $\nu_a(E) = \nu(A \cap E)$, $\nu_s(E) = \nu(B \cap E)$. Then $\nu_s \perp \mu$ and $d\nu_a = g(1-g)^{-1}\chi_A\, d\mu$.

19. Define $\varphi_n \in (l^\infty)^*$ by $\varphi_n(f) = n^{-1}\sum_1^n f(j)$. Then the sequence $\{\varphi_n\}$ has a weak * cluster point φ, and φ is an element of $(l^\infty)^*$ which does not arise from an element of l^1.

20. Suppose $\sup_n \|f_n\|_p < \infty$ and $f_n \to f$ a.e.

(a) If $1 < p < \infty$, then $f_n \to f$ weakly in L^p. [Given $g \in L^q$, where q is conjugate to p, and $\varepsilon > 0$, there exist (i) $\delta > 0$ such that $\int_E |g|^q < \varepsilon$ whenever $\mu(E) < \delta$, (ii) $A \subset X$ such that $\mu(A) < \infty$ and $\int_{A^c} |g|^q < \varepsilon$, and (iii) $B \subset A$ such that $\mu(B \setminus A) < \delta$ and $f_n \to f$ uniformly on B.]

(b) The result of (a) is false in general for $p = 1$. [Find counterexamples in $L^1(\mathbf{R}, m)$ and l^1.] It is, however, true for $p = \infty$ if μ is σ-finite and weak convergence is replaced by weak * convergence.

21. If $1 < p < \infty$, $f_n \to f$ weakly in $l^p(A)$ iff $\sup_n \|f_n\|_p < \infty$ and $f_n \to f$ pointwise.

22. Let $X = [0, 1]$, with Lebesgue measure.

(a) Let $f_n(x) = \cos 2\pi nx$. Then $f_n \to 0$ weakly in L^2 (cf. Exercise 5.63), but $f_n \not\to 0$ a.e. or in measure.

(b) Let $f_n(x) = n\chi_{(0,1/n)}$. Then $f_n \to 0$ a.e. and in measure, but $f_n \not\to 0$ weakly in L^p for any p.

The next four exercises deal with a modification of L^∞ which is useful in non-semifinite measure spaces.

23. Let $(X, \mathscr{M}, \mu)$ be a measure space. A set $E \in \mathscr{M}$ is called **locally null** if $\mu(E \cap F) = 0$ for every $F \in \mathscr{M}$ such that $\mu(F) < \infty$. If E is locally null, then $\mu(E)$ is either 0 or ∞. If μ is semifinite, then every locally null set is null.

24. Let $\mathscr{L}^\infty = \mathscr{L}^\infty(X, \mathscr{M}, \mu)$ be the space of all measurable $f: X \to \mathbf{C}$ such that $\|f\|_* < \infty$, where

$$\|f\|_* = \inf\{a: \{x: |f(x)| > a\} \text{ is locally null}\}.$$

We consider $f, g \in \mathscr{L}^\infty$ to be identical if $\{x: f(x) \neq g(x)\}$ is locally null. Then $\|\ \|_*$ is a norm on $\mathscr{L}^\infty$ which makes $\mathscr{L}^\infty$ into a Banach space.

25. If $g \in \mathscr{L}^\infty$, then $\|g\|_* = \sup\{|\int fg|: \|f\|_1 = 1\}$, so the map $f \to \varphi_f$ is an isometry from $\mathscr{L}^\infty$ into $(L^1)^*$. Conversely, if g is measurable, $fg \in L^1$ for all $f \in S$ [where S is as in Theorem (6.14)], and $\sup\{|\int fg|: f \in S, \|f\|_1 = 1\} < \infty$, then $g \in \mathscr{L}^\infty$.

26. Suppose μ is decomposable (Exercise 3.15). Then every $\varphi \in (L^1)^*$ is of the form $\varphi(f) = \int fg$ for some $g \in \mathscr{L}^\infty$, and hence $(L^1)^* \cong \mathscr{L}^\infty$. (If $\mathscr{F}$ is a decomposition of μ and $f \in L^1$, there exists $\{E_j\} \subset \mathscr{F}$ such that $f = \sum_1^\infty f\chi_{E_j}$ where the series converges in L^1.)

3. SOME USEFUL INEQUALITIES

Estimates and inequalities lie at the heart of the applications of L^p spaces in analysis. The most basic of these are the Hölder and Minkowski inequalities; in this section we present a few more important results in this area. The first one is almost a triviality, but is sufficiently useful to warrant special mention.

(6.17) Chebyshev's Inequality. If $f \in L^p$ $(0 < p < \infty)$, then for any $\alpha > 0$,

$$\mu(\{x: |f(x)| > \alpha\}) \le \left(\frac{\|f\|_p}{\alpha}\right)^p.$$

Proof. Let $E_\alpha = \{x: |f(x)| > \alpha\}$. Then

$$\|f\|_p^p = \int |f|^p \ge \int_{E_\alpha} |f|^p \ge \alpha^p \int_{E_\alpha} 1 = \alpha^p \mu(E_\alpha). \qquad \square$$

The next result is a rather general theorem about boundedness of integral operators on L^p spaces.

(6.18) Theorem. Let $(X, \mathscr{M}, \mu)$ and $(Y, \mathscr{N}, \nu)$ be σ-finite measure spaces, and let K be an $\mathscr{M} \otimes \mathscr{N}$-measurable function on $X \times Y$. Suppose there exists $C > 0$ such that $\int |K(x, y)|\, d\mu(x) \le C$ for a.e. $y \in Y$ and $\int |K(x, y)|\, d\nu(y) \le C$ for a.e. $x \in X$. If $f \in L^p(\nu)$ $(1 \le p \le \infty)$, the integral

$$Tf(x) = \int K(x, y) f(y)\, d\nu(y)$$

converges absolutely for a.e. $x \in X$, the function Tf thus defined is in $L^p(\mu)$, and $\|Tf\|_p \le C\|f\|_p$.

Proof. Suppose $1 < p < \infty$. If q is the conjugate exponent to p, by Hölder's inequality we have

$$\begin{aligned}\int|K(x,y)f(y)|\,d\nu(y) &= \int|K(x,y)|^{(1/q)+(1/p)}|f(y)|\,d\nu(y)\\ &\le \left[\int|K(x,y)|\,d\nu(y)\right]^{1/q}\left[\int|K(x,y)||f(y)|^p\,d\nu(y)\right]^{1/p}\\ &\le C^{1/q}\left[\int|K(x,y)||f(y)|^p\,d\nu(y)\right]^{1/p}\end{aligned}$$

for a.e. $x \in X$. Hence, by Tonelli's theorem,

$$\begin{aligned}&\int\left[\int|K(x,y)f(y)|\,d\nu(y)\right]^p d\mu(x)\\ &\quad\le C^{p/q}\iint|K(x,y)||f(y)|^p\,d\nu(y)\,d\mu(x)\\ &\quad\le C^{(p/q)+1}\int|f(y)|^p\,d\nu(y).\end{aligned}$$

Since the last integral is finite, Fubini's theorem implies that $K(x,\cdot)f \in L^1(\nu)$ for a.e. x, so that Tf is well defined a.e., and

$$\int|Tf(x)|^p\,d\mu(x) \le C^{(p/q)+1}\|f\|_p^p.$$

Taking pth roots, we are done.

For $p = 1$, the proof is similar but easier and requires only the hypothesis $\int|K(x,y)|\,d\mu(x) \le C$; for $p = \infty$ the proof is trivial and requires only the hypothesis $\int|K(x,y)|\,d\nu(y) \le C$. □

Minkowski's inequality states that the L^p norm of a sum is at most the sum of the L^p norms. There is a generalization of this result in which sums are replaced by integrals:

(6.19) Minkowski's Inequality for Integrals. Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces, and let f be an $\mathcal{M}\otimes\mathcal{N}$-measurable function on $X \times Y$.

(a) If $f \ge 0$ and $1 \le p < \infty$, then

$$\left[\int\left(\int f(x,y)\,d\nu(y)\right)^p d\mu(x)\right]^{1/p} \le \int\left[\int f(x,y)^p\,d\mu(x)\right]^{1/p} d\nu(y).$$

(b) If $1 \le p \le \infty$, $f(\cdot,y) \in L^p(\mu)$ for a.e. y, and the function $y \to \|f(\cdot,y)\|_p$ is in $L^1(\nu)$, then $f(x,\cdot) \in L^1(\nu)$ for a.e. x, the function $x \to \int f(x,y)\,d\nu(y)$ is in $L^p(\mu)$, and

$$\left\|\int f(\cdot,y)\,d\nu(y)\right\|_p \le \int\|f(\cdot,y)\|_p\,d\nu(y).$$

Proof. (a) is merely Tonelli's theorem when $p = 1$. If $1 < p < \infty$, let q be the conjugate exponent to p and suppose $g \in L^q(\mu)$. Then by Tonelli's theorem and Hölder's inequality,

$$\begin{aligned}\int\left[\int f(x,y)\,d\nu(y)\right]|g(x)|\,d\mu(x) &= \iint f(x,y)|g(x)|\,d\mu(x)\,d\nu(y)\\ &\le \int\left[\int f(x,y)^p\,d\mu(x)\right]^{1/p} d\nu(y)\,\|g\|_q.\end{aligned}$$

Assertion (a) therefore follows from the converse to Holder's inequality. When $p < \infty$, (b) follows from (a) (with f replaced by $|f|$) and Fubini's theorem; when $p = \infty$ it is a simple consequence of the monotonicity of the integral. □

Our final result is a theorem of Hardy, Littlewood, and Pólya concerning integral operators on $(0,\infty)$ with Lebesgue measure.

(6.20) Theorem. Suppose K is a Lebesgue measurable function on $(0,\infty)\times(0,\infty)$ such that (i) $K(\lambda x, \lambda y) = \lambda^{-1}K(x,y)$ for all $\lambda > 0$, and (ii) $\int_0^\infty|K(x,1)|x^{-1/p}\,dx = C < \infty$, where $1 \le p \le \infty$. Suppose also that q is the conjugate exponent to p. If $f \in L^p$ and $g \in L^q$, let

$$Tf(y) = \int_0^\infty K(x,y)f(x)\,dx, \quad Sg(x) = \int_0^\infty K(x,y)g(y)\,dy.$$

Then Tf and Tg are defined a.e., and $\|Tf\|_p \le C\|f\|_p$ and $\|Sg\|_q \le C\|g\|_q$.

Proof. Setting $z = x/y$,

$$\int_0^\infty|K(x,y)f(x)|\,dx = \int_0^\infty|K(yz,y)f(yz)|\,y\,dz = \int_0^\infty|K(z,1)f_z(y)|\,dz$$

where $f_z(y) = f(yz)$; moreover,

$$\|f_z\|_p = \left[\int_0^\infty|f(yz)|^p\,dy\right]^{1/p} = \left[\int_0^\infty z^{-1}|f(x)|^p\,dx\right]^{1/p} = z^{-1/p}\|f\|_p.$$

Therefore, by Minkowski's inequality for integrals, Tf exists a.e., and

$$\|Tf\|_p \le \int_0^\infty|K(z,1)|\,\|f_z\|_p\,dz \le \|f\|_p\int_0^\infty|K(z,1)|z^{-1/p}\,dx \le C\|f\|_p.$$

Finally, setting $u = 1/y$,

$$\begin{aligned}\int_0^\infty|K(1,y)|y^{-1/q}\,dy &= \int_0^\infty|K(y^{-1},1)|y^{-1-(1/q)}\,dy\\ &= \int_0^\infty|K(u,1)|u^{-1/p}\,du = C,\end{aligned}$$

so the same reasoning shows that $\|Sg\|_q \le \|g\|_q$. □

(6.21) Corollary (Hardy's Inequalities). Suppose $1 < p \le \infty$ and q is the conjugate exponent to p. Let

$$Tf(x) = x^{-1}\int_0^x f(y)\,dy, \quad Sg(x) = \int_x^\infty y^{-1}g(y)\,dy.$$

Then

$$\|Tf\|_p \le \frac{p}{p-1}\|f\|_p, \quad \|Sg\|_q \le \frac{p}{p-1}\|g\|_q.$$

Proof. Take $K(x, y) = y^{-1}\chi_E(x, y)$ where $E = \{(x, y): x < y\}$ in Theorem (6.20). □

EXERCISES

27. (Hilbert's Inequality) The operator $Tf(x) = \int_0^\infty (x + y)^{-1}f(y)\,dy$ satisfies $\|Tf\|_p \le C_p\|f\|_p$ for $1 < p < \infty$, where $C_p = \int_0^\infty x^{-1/p}(x + 1)^{-1}dx$. [For those who know about contour integrals: show that $C_p = \pi\csc(\pi/p)$.]

28. Let I_α be the αth fractional integration operator as in Exercise 2.63, and let $J_\alpha f(x) = x^{-\alpha}I_\alpha f(x)$.

(a) J_α is bounded on $L^p(0, \infty)$ for $1 < p \le \infty$; more precisely,

$$\|J_\alpha f\|_p \le \frac{\Gamma(1 - p^{-1})}{\Gamma(\alpha + 1 - p^{-1})}\|f\|_p.$$

(b) Find an $f \in L^1(0, \infty)$ such that $J_1 f \notin L^1(0, \infty)$.

29. (More Hardy inequalities) Suppose $1 \le p < \infty$, $r > 0$, and h is a nonnegative measurable function on $(0, \infty)$. Then:

$$\int_0^\infty x^{-r-1}\left[\int_0^x h(y)\,dy\right]^p dx \le \left(\frac{p}{r}\right)^p \int_0^\infty x^{p-r-1}h(x)^p\,dx,$$

$$\int_0^\infty x^{r-1}\left[\int_x^\infty h(y)\,dy\right]^p dx \le \left(\frac{p}{r}\right)^p \int_0^\infty x^{p+r-1}h(x)^p\,dx.$$

[Apply Theorem (6.20) with $K(x, y) = x^{\beta-1}y^{-\beta}$ for $x \le y$ and $K(x, y) = 0$ for $x > y$, $f(x) = x^\gamma h(x)$, and $g(x) = x^\delta h(x)$ for suitable β, γ, δ.]

30. Suppose K is a nonnegative measurable function on $(0, \infty)$ such that $\int_0^\infty K(x)x^{s-1}\,dx = \varphi(s) < \infty$ for $0 < s < 1$.

(a) If $1 < p < \infty$, $p^{-1} + q^{-1} = 1$, and f, g are nonnegative measurable functions on $(0, \infty)$,

$$\int_0^\infty\int_0^\infty K(xy)f(x)g(y)\,dx\,dy \le \varphi(p^{-1})\left[\int_0^\infty x^{p-2}f(x)^p dx\right]^{1/p}\left[\int_0^\infty g(x)^q dx\right]^{1/q}.$$

(b) The operator $Tf(x) = \int_0^\infty K(xy)f(y)\,dy$ is bounded on $L^2(0, \infty)$. [Interesting special case: if $K(x) = e^{-x}$, then T is the Laplace transform and $\varphi(s) = \Gamma(s)$.]

31. (Generalized Hölder inequality) Suppose $1 \le p_j \le \infty$ and $\sum_1^n p_j^{-1} = r^{-1} \le 1$. If $f_j \in L^{p_j}$ for $j = 1, \dots, n$, the $\prod_1^n f_j \in L^r$ and $\|\prod_1^n f_j\|_r \le \prod_1^n \|f_j\|_{p_j}$. (First do the case $n = 2$.)

32. (Heisenberg's inequality) Suppose f is absolutely continuous on $[-a, a]$ for all $a \in \mathbf{R}$, so that f' exists a.e. on $\mathbf{R}$. Then

$$\int_{-\infty}^\infty |f(x)|^2\,dx \le 2\left[\int_{-\infty}^\infty |xf(x)|^2\,dx\right]^{1/2}\left[\int_{-\infty}^\infty |f'(x)|^2\,dx\right]^{1/2}.$$

(If the integrals on the right are finite, one can integrate by parts to obtain $\int|f|^2 = -2\,\mathrm{Re}\int x\bar{f}f'$. The reader who knows quantum mechanics will recognize this inequality as the position–momentum uncertainty principle.)

33. If $1 < p < \infty$, $p^{-1} + q^{-1} = 1$, $f \in L^p(0, \infty)$, and $F(x) = \int_0^x f(t)\,dt$, then $x^{-1/q}F(x) \to 0$ as $x \to 0$ and as $x \to \infty$.

34. If f is absolutely continuous on $[\varepsilon, 1]$ for $0 < \varepsilon < 1$ and $\int_0^1 x|f'(x)|^p\,dx < \infty$, then $\lim_{x\to 0} f(x)$ exists (and is finite) if $p > 2$, $|f(x)|/|\log x|^{1/2} \to 0$ as $x \to 0$ if $p = 2$, and $|f(x)|/x^{1-(2/p)} \to 0$ as $x \to 0$ if $p < 2$.

4. DISTRIBUTION FUNCTIONS AND WEAK L^p

If f is a measurable function on $(X, \mathcal{M}, \mu)$, we define its **distribution function** $\lambda_f: (0, \infty) \to [0, \infty]$ by

$$\lambda_f(\alpha) = \mu(\{x: |f(x)| > \alpha\}).$$

[This is closely related, but not identical, to the "distribution function" used in probability theory, which is useful only when $\mu(X) < \infty$. See Section 9.1.] We compile the basic properties of λ_f in a proposition:

(6.22) Proposition.

(a) λ_f is decreasing and right continuous.

(b) If $|f| \le |g|$, then $\lambda_f \le \lambda_g$.

(c) If $|f_n|$ increases to $|f|$, then λ_{f_n} increases to λ_f.

(d) If $f = g + h$, then $\lambda_f(\alpha) \le \lambda_g(\alpha/2) + \lambda_h(\alpha/2)$.

Proof. Let $E(\alpha, f) = \{x: |f(x)| > \alpha\}$. λ_f is decreasing since $E(\alpha, f) \supset E(\beta, f)$ if $\alpha > \beta$, and it is right continuous since $E(\alpha, f)$ is the increasing union of $\{E(\alpha + \varepsilon_n, f)\}$ for any sequence $\{\varepsilon_n\}$ which decreases to zero. If $|f| \le |g|$, then $E(\alpha, f) \subset E(\alpha, g)$, so $\lambda_f \le \lambda_g$. If $|f_n|$ increases to $|f|$, then $E(\alpha, f)$ is the increasing union of $\{E(\alpha, f_n)\}$, so λ_{f_n} increases to λ_f. Finally, if

$f = g + h$, then $E(\alpha, f) \subset E(\alpha/2, g) \cup E(\alpha/2, h)$, which implies that $\lambda_f(\alpha) \leq \lambda_g(\alpha/2) + \lambda_h(\alpha/2)$. □

Suppose that $\lambda_f(\alpha) < \infty$ for all $\alpha > 0$. In view of Proposition (6.22a), λ_f defines a negative Borel measure ν on $(0, \infty)$ such that $\nu((a, b]) = \lambda_f(b) - \lambda_f(a)$ for all $a, b > 0$. [Our construction of Borel measures on $\mathbf{R}$ in Section 1.5 works equally well on $(0, \infty)$.] We can therefore consider the Lebesgue–Stieltjes integrals $\int \varphi \, d\lambda_f = \int \varphi \, d\nu$ of functions φ on $(0, \infty)$. The following result shows that integrals of functions of $|f|$ on X can be reduced to such Lebesgue–Stieltjes integrals.

(6.23) Proposition. If $\lambda_f(\alpha) < \infty$ for all $\alpha > 0$ and φ is a nonnegative Borel measurable function on $(0, \infty)$, then

$$\int_X \varphi \circ |f| \, d\mu = -\int_0^\infty \varphi(\alpha) \, d\lambda_f(\alpha).$$

Proof. If ν is the negative measure determined by λ_f, we have

$$\nu((a, b]) = \lambda_f(b) - \lambda_f(a) = -\mu(\{x: a < |f(x)| \leq b\})$$

$$= -\mu(|f|^{-1}((a, b])).$$

By the uniqueness of extensions [Theorem (1.14)], it follows that $\nu(E) = -\mu(|f|^{-1}(E))$ for all Borel sets $E \subset (0, \infty)$. But this means that $\int_X \varphi \circ |f| \, d\mu = -\int_0^\infty \varphi(\alpha) \, d\lambda_f(\alpha)$ when φ is the characteristic function of a Borel set, and hence when φ is simple. The general case then follows by virtue of Theorem (2.10) and the monotone convergence theorem. □

The case of this result in which we are most interested is $\varphi(\alpha) = \alpha^p$, which gives

$$\int |f|^p \, d\mu = -\int_0^\infty \alpha^p \, d\lambda_f(\alpha).$$

A more useful form of this equation is obtained by integrating the right side by parts [Theorem (3.30)] to obtain

$$\int |f|^p \, d\mu = p \int_0^\infty \alpha^{p-1} \lambda_f(\alpha) \, d\alpha.$$

The validity of this calculation is not clear unless we know that $\alpha^p \lambda_f(\alpha) \to 0$ as $\alpha \to 0$ and $\alpha \to \infty$; nonetheless, the conclusion is correct.

(6.24) Proposition. If $0 < p < \infty$, then $\int |f|^p \, d\mu = p \int_0^\infty \alpha^{p-1} \lambda_f(\alpha) \, d\alpha$.

Proof. If $\lambda_f(\alpha) = \infty$ for some $\alpha > 0$, then both integrals are infinite. If not, and f is simple, then λ_f is bounded as $\alpha \to 0$ and vanishes for α sufficiently large, so the integration by parts described above works. (It is also easy to verify the formula directly in this case.) For the general case, let $\{g_n\}$ be a sequence of simple functions which increase to $|f|$; then the desired result is true for g_n, and it follows for f by Proposition (6.22c) and the monotone convergence theorem. □

A variant of the L^p spaces which turns up rather often is the following. If f is a measurable function on X and $0 < p < \infty$, we define

$$[f]_p = \left(\sup_{\alpha > 0} \alpha^p \lambda_f(\alpha)\right)^{1/p},$$

and we define **weak L^p** to be the set of all f such that $[f]_p < \infty$. $[\,]_p$ is not a norm: it is easy to check that $[cf]_p = |c|[f]_p$, but the triangle inequality fails. However, weak L^p is a topological vector space: see Exercise 35.

The relationship between L^p and weak L^p is the following: On the one hand,

$$L^p \subset \text{weak } L^p \quad \text{and} \quad [f]_p \leq \|f\|_p.$$

(This is just a restatement of Chebyshev's inequality.) On the other hand, if we majorize $\lambda_f(\alpha)$ by $([f]_p/\alpha)^p$ in the integral $p \int_0^\infty \alpha^{p-1} \lambda_f(\alpha) \, d\alpha$, which equals $\|f\|_p^p$, we obtain a constant times $\int_0^\infty \alpha^{-1} \, d\alpha$, which is divergent at both 0 and ∞—but just barely. One needs only slightly stronger estimates on λ_f near 0 and ∞ to obtain $f \in L^p$. (See also Exercise 36.) The standard example of a function which is in weak L^p but not in L^p is $f(x) = x^{-1/p}$ on $(0, \infty)$ (with Lebesgue measure).

Frequently it is convenient to express a function as the sum of a "small" part and a "big" part. The following is a way of doing this which gives a simple formula for the distribution functions.

(6.25) Proposition. If f is a measurable function and $A > 0$, let $E(A) = \{x: |f(x)| > A\}$, and set

$$h_A = f\chi_{E(A)^c} + A(\operatorname{sgn} f)\chi_{E(A)}, \quad g_A = f - h_A = (\operatorname{sgn} f)(|f| - A)\chi_{E(A)}.$$

Then $\lambda_{h_A}(\alpha) = \lambda_f(\alpha)$ if $\alpha < A$, $\lambda_{h_A}(\alpha) = 0$ if $\alpha \geq A$, and $\lambda_g(\alpha) = \lambda_f(\alpha + A)$.

The proof is left to the reader (Exercise 37).

EXERCISES

35. For any measurable f and g we have $[cf]_p = |c|[f]_p$ and $[f+g]_p \le 2([f]_p^p + [g]_p^p)^{1/p}$; hence weak L^p is a vector space. Moreover, the "balls" $\{g \in \text{weak } L^p: [g-f]_p < r\}$ $(r > 0, f \in \text{weak } L^p)$ generate a topology on weak L^p which makes weak L^p into a topological vector space.

36. If $f \in$ weak L^p and $\mu(\{x: f(x) \neq 0\}) < \infty$, then $f \in L^q$ for all $q < p$. On the other hand, if $f \in (\text{weak } L^p) \cap L^\infty$, then $f \in L^q$ for all $q > p$.

37. Prove Proposition (6.25).

38. $f \in L^p$ iff $\sum_{-\infty}^{\infty} 2^{kp}\lambda_f(2^k) < \infty$.

39. If $f \in L^p$, then $\lim_{\alpha\to 0}\alpha^p\lambda_f(\alpha) = \lim_{\alpha\to\infty}\alpha^p\lambda_f(\alpha) = 0$. (First suppose f is simple.)

40. If f is a measurable function on X, its **decreasing rearrangement** is the function $f^*:(0,\infty) \to [0,\infty)$ defined by

$$f^*(t) = \inf\{\alpha: \lambda_f(\alpha) \le t\}.$$

(a) f^* is decreasing, $\lambda_f(f^*(t)) \le t$ for all $t > 0$, and $f^*(\lambda_f(\alpha)) \le \alpha$ for all $\alpha > 0$.

(b) $\lambda_f = \lambda_{f^*}$, where λ_{f^*} is defined with respect to Lebesgue measure on $(0,\infty)$.

(c) If $\lambda_f(\alpha) < \infty$ for all $\alpha > 0$ and φ is a nonnegative Borel measurable function on $(0,\infty)$, then

$$\int \varphi \circ |f|\, d\mu = \int_0^\infty \varphi \circ f^*(t)\, dt.$$

In particular, $\|f\|_p = \|f^*\|_p$ for $0 < p < \infty$.

(d) If $0 < p < \infty$, $[f]_p = \sup_{t>0} t^{1/p} f^*(t)$.

(e) The name "rearrangement" for f^* derives from the case where f is a nonnegative function on $(0,\infty)$. To see why it is appropriate, pick a step function f on $(0,\infty)$ assuming four or five different values and draw the graphs of f and f^*.

5. INTERPOLATION OF L^p SPACES

If $p < q < r$, then $(L^p \cap L^r) \subset L^q \subset (L^p + L^r)$, and it is natural to ask whether a linear operator T on $L^p + L^r$ which is bounded on both L^p and L^r is also bounded on L^q. The answer is affirmative, and this result can be generalized in various ways. The two fundamental theorems on this question are the Riesz–Thorin and Marcinkiewicz interpolation theorems, which we

present in this section. We begin with the Riesz–Thorin theorem, whose proof is based on the following lemma from complex function theory.

(6.26) The Three Lines Lemma. Let φ be a bounded continuous function on the strip $0 \le \operatorname{Re} z \le 1$ which is holomorphic on the interior of the strip. If $|\varphi(z)| \le M_0$ for $\operatorname{Re} z = 0$ and $|\varphi(z)| \le M_1$ for $\operatorname{Re} z = 1$, then $|\varphi(z)| \le M_0^{1-t}M_1^t$ for $\operatorname{Re} z = t$, $0 < t < 1$.

Proof. For $\varepsilon > 0$ let $\varphi_\varepsilon(z) = \varphi(z)M_0^{z-1}M_1^{-z}\exp(\varepsilon(z^2-1))$. Then φ_ε satisfies the hypotheses of the lemma with M_0, M_1 replaced by 1, and also $|\varphi_\varepsilon(z)| \to 0$ as $|\operatorname{Im} z| \to \infty$. Thus $|\varphi_\varepsilon(z)| \le 1$ on the boundary of the rectangle $0 \le \operatorname{Re} z \le 1$, $-A \le \operatorname{Im} z \le A$ provided A is large, and the maximum modulus principle therefore implies that $|\varphi_\varepsilon(z)| \le 1$ in the strip $0 \le \operatorname{Re} z \le 1$. Letting $\varepsilon \to 0$ we obtain the desired result:

$$|\varphi(z)|M_0^{t-1}M_1^{-t} = \lim_{\varepsilon\to 0}|\varphi_\varepsilon(z)| \le 1 \text{ for } \operatorname{Re} z = t. \qquad \square$$

(6.27) The Riesz–Thorin Interpolation Theorem. Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are semifinite measure spaces and p_0, p_1, q_0, q_1 are elements of $[1,\infty]$, and for $0 < t < 1$ let p_t, q_t be defined by

$$\frac{1}{p_t} = \frac{1-t}{p_0} + \frac{t}{p_1}, \qquad \frac{1}{q_t} = \frac{1-t}{q_0} + \frac{t}{q_1}.$$

If T is a linear map from $L^{p_0}(\mu) + L^{p_1}(\mu)$ into $L^{q_0}(\nu) + L^{q_1}(\nu)$ such that $\|Tf\|_{q_0} \le M_0\|f\|_{p_0}$ for $f \in L^{p_0}(\mu)$ and $\|Tf\|_{q_1} \le M_1\|f\|_{p_1}$ for $f \in L^{p_1}(\mu)$, then $\|Tf\|_{q_t} \le M_0^{1-t}M_1^t\|f\|_{p_t}$ for $f \in L^{p_t}(\mu)$, $0 < t < 1$.

Proof. To begin with, we observe that the case $p_0 = p_1$ follows from Proposition (6.10): if $p = p_0 = p_1$, then

$$\|Tf\|_{q_t} \le \|Tf\|_{q_0}^{1-t}\|Tf\|_{q_1}^t \le M_0^{1-t}M_1^t\|f\|_p.$$

Thus we may assume that $p_0 \neq p_1$, and in particular that $p_t < \infty$ for $0 < t < 1$.

Let S_X (resp. S_Y) be the space of all simple functions on X (resp. Y) which vanish outside sets of finite measure. Then $S_X \subset L^p(\mu)$ for all p and S_X is dense in $L^p(\mu)$ for $p < \infty$; similarly for S_Y. The main part of the proof consists in showing that $\|Tf\|_{q_t} \le M_0^{1-t}M_1^t\|f\|_{p_t}$ for all $f \in S_X$. However, by the converse to Hölder's inequality,

$$\|Tf\|_{q_t} = \sup\left\{\left|\int (Tf)g\, d\nu\right|: g \in S_Y \text{ and } \|g\|_{q_t'} = 1\right\}$$

where q_t' is the conjugate exponent to q_t. Moreover, we may assume that $f \neq 0$ and rescale f so that $\|f\|_{p_t} = 1$. We therefore wish to establish the following claim: If $f \in S_X$ and $\|f\|_{p_t} = 1$, then $|\int (Tf)g\, d\nu| \le M_0^{1-t}M_1^t$ for all $g \in S_Y$ such that $\|g\|_{q_t'} = 1$.

Let $f = \sum_1^m a_j \chi_{E_j}$ and $g = \sum_1^n b_k \chi_{F_k}$ where the E_j's and the F_j's are disjoint in X and Y, and let $a_j = |a_j| e^{i\theta_j}$, $b_k = |b_k| e^{i\psi_k}$. Moreover, let $\alpha_j = p_j^{-1}$ and $\beta_j = q_j^{-1}$ for $j = 0, 1$ and define $\alpha(z) = (1 - z)\alpha_0 + z\alpha_1$, $\beta(z) = (1 - z)\beta_0 + z\beta_1$ for $z \in \mathbf{C}$; then $\alpha(t) = p_t^{-1}$ and $\beta(t) = q_t^{-1}$. Fix $t \in (0, 1)$: we have assumed that $p_t < \infty$ and hence $\alpha_t > 0$, so we may define

$$f_z = \sum_1^m |a_j|^{\alpha(z)/\alpha(t)} e^{i\theta_j} \chi_{E_j}.$$

If $\beta(t) < 1$, we define

$$g_z = \sum_1^n |b_k|^{(1-\beta(z))/(1-\beta(t))} e^{i\psi_k} \chi_{F_k},$$

while if $\beta(t) = 1$ we define $g_z = g$ for all z. [We shall assume henceforth that $\beta(t) < 1$ and leave the easy modification for $\beta(t) = 1$ to the reader.] Finally, we set $\varphi(z) = \int (Tf_z) g_z \, d\nu$. Thus

$$\varphi(z) = \sum_{j,k} |a_j|^{\alpha(z)/\alpha(t)} |b_k|^{(1-\beta(z))/(1-\beta(t))} c_{jk},$$

$$\text{where } c_{jk} = e^{i\theta_j + i\psi_k} \int \left(T\chi_{E_j}\right) \chi_{F_k} \, d\nu,$$

so that $\varphi(z)$ is an entire holomorphic function of z which is bounded in the strip $0 \le \operatorname{Re} z \le 1$. Since $\varphi(t) = \int (Tf) g \, d\nu$, by the three lines lemma it will suffice to show that $|\varphi(z)| \le M_0$ for $\operatorname{Re} z = 0$ and $|\varphi(z)| \le M_1$ for $\operatorname{Re} z = 1$. However, since for $s \in \mathbf{R}$ we have $\alpha(is) = \alpha_0 + is(\alpha_1 - \alpha_0)$ and $1 - \beta(is) = (1 - \beta_0) - is(\beta_1 - \beta_0)$,

$$|f_{is}| = |f|^{\alpha_0/\alpha(t)} = |f|^{p_t/p_0},$$

$$|g_{is}| = |g|^{(1-\beta_0)/(1-\beta(t))} = |g|^{q_t'/q_0'}.$$

Therefore, by Hölder's inequality,

$$|\varphi(is)| \le \|Tf_{is}\|_{q_0} \|g_{is}\|_{q_0'}$$

$$\le M_0 \|f_{is}\|_{p_0} \|g_{is}\|_{q_0'} = M_0 \|f\|_{p_t} \|g\|_{q_t'} = M_0.$$

A similar calculation shows that $|\varphi(1 + is)| \le M_1$, so the claim is proved.

We have shown that $\|Tf\|_{q_t} \le M_0^{1-t} M_1^t \|f\|_{p_t}$ for $f \in S_X$, so $T|S_X$ has a unique extension to $L^{p_t}(\mu)$ satisfying the same estimates there. It remains to show that this extension is T itself, that is, that T satisfies this estimate for all $f \in L^{p_t}(\mu)$. Given such an f, choose a sequence $\{f_n\}$ in S_X such that $|f_n| \le |f|$

and $f_n \to f$ pointwise; also, if $E = \{x: |f(x)| > 1\}$, let $g = f\chi_E$, $g_n = f_n \chi_E$, $h = f - g$, and $h_n = f_n - g_n$. Then if $p_0 < p_1$ (which we may assume, by relabeling) we have $g \in L^{p_0}(\mu)$, $h \in L^{p_1}(\mu)$, and by the dominated convergence theorem, $\|f_n - f\|_{p_t} \to 0$, $\|g_n - g\|_{p_0} \to 0$, $\|h_n - h\|_{p_1} \to 0$. Hence $\|Tg_n - Tg\|_{q_0} \to 0$ and $\|Th_n - Th\|_{q_1} \to 0$, so by passing to a suitable subsequence we may assume that $Tg_n \to Tg$ and $Th_n \to Th$ a.e. (see Exercise 9). But then $Tf_n \to Tf$ a.e., so by Fatou's lemma,

$$\|Tf\|_{q_t} \le \liminf \|Tf_n\|_{q_t} \le \liminf M_0^{1-t} M_1^t \|f_n\|_{p_t} = M_0^{1-t} M_1^t \|f\|_{p_t}. \qquad \square$$

The conclusion of the Riesz–Thorin theorem can be restated in a slightly stronger form. Let $M(t)$ be the operator norm of T as a map from $L^{p_t}(\mu)$ to $L^{q_t}(\nu)$. We have shown that $M(t) \le M(0)^{1-t} M(1)^t$. It is possible for strict inequality to hold; however, if $0 < s < t < u < 1$ and $t = (1 - \tau)s + \tau u$, the theorem may be applied again to show that $M(t) \le M(s)^{1-\tau} M(u)^\tau$. In short, the conclusion is that $\log M(t)$ is a convex function on $[0, 1]$.

We now turn to the Marcinkiewicz theorem, for which we need some more terminology. Let T be a map from some vector space $\mathscr{D}$ of measurable functions on $(X, \mathscr{M}, \mu)$ to the space of all measurable functions on $(Y, \mathscr{N}, \nu)$. T is called **sublinear** if $|T(f + g)| \le |Tf| + |Tg|$ and $|T(cf)| = |c||Tf|$ for all $f, g \in \mathscr{D}$ and $c \in \mathbf{C}$. (For example, every linear map is sublinear.) A sublinear map T is said to be **strong type** (p, q) $(1 \le p, q \le \infty)$ if $L^p(\mu) \subset \mathscr{D}$, T maps $L^p(\mu)$ into $L^q(\nu)$, and there exists $C > 0$ such that $\|Tf\|_q \le C\|f\|_p$ for all $f \in L^p(\mu)$. T is said to be **weak type** (p, q) $(1 \le p \le \infty, 1 \le q < \infty)$ if $L^p(\mu) \subset \mathscr{D}$, T maps $L^p(\mu)$ into weak $L^q(\nu)$, and there exists $C > 0$ such that $[Tf]_q \le C\|f\|_p$ for all $f \in L^p(\mu)$. Finally, T is said to be **weak type** (p, ∞) if T is strong type (p, ∞).

(6.28) The Marcinkiewicz Interpolation Theorem. Suppose $(X, \mathscr{M}, \mu)$ and $(Y, \mathscr{N}, \nu)$ are measure spaces; p_0, p_1, q_0, q_1 are elements of $[1, \infty]$ such that $p_0 \le q_0$, $p_1 \le q_1$, and $q_0 \ne q_1$; and

$$\frac{1}{p} = \frac{1 - t}{p_0} + \frac{t}{p_1}, \quad \frac{1}{q} = \frac{1 - t}{q_0} + \frac{t}{q_1}, \quad \text{where } 0 < t < 1.$$

If T is a sublinear map from $L^{p_0}(\mu) + L^{p_1}(\mu)$ to the space of measurable functions on Y which is weak types (p_0, q_0) and (p_1, q_1), then T is strong type (p, q). More precisely, if $[Tf]_{q_j} \le C_j \|f\|_{p_j}$ for $j = 0, 1$, then $\|Tf\|_q \le B_p \|f\|_p$ where B_p depends only on p_j, q_j, C_j in addition to p; and for $j = 0, 1$, $B_p|p - p_j|$ (resp. B_p) remains bounded as $p \to p_j$ if $p_j < \infty$ (resp. $p_j = \infty$).

Proof. The case $p_0 = p_1$ is easy and is left to the reader (Exercise 42). Without loss of generality we therefore assume that $p_0 < p_1$, and for the time being we also assume that $q_0 < \infty$ and $q_1 < \infty$ (whence also $p_0 < p_1 < \infty$). Given $f \in L^p(\mu)$ and $A > 0$, let g_A, h_A be as in Proposition (6.25). Then by

Propositions (6.24) and (6.25),

$$\text{(6.29a)}\quad \int |g_A|^{p_0}\,d\mu = p_0\int_0^\infty \beta^{p_0-1}\lambda_{g_A}(\beta)\,d\beta = p_0\int_0^\infty \beta^{p_0-1}\lambda_f(\beta+A)\,d\beta$$

$$= p_0\int_A^\infty (\beta-A)^{p_0-1}\lambda_f(\beta)\,d\beta \le p_0\int_A^\infty \beta^{p_0-1}\lambda_f(\beta)\,d\beta,$$

$$\text{(6.29b)}\quad \int |h_A|^{p_1}\,d\mu = p_1\int_0^\infty \beta^{p_1-1}\lambda_{h_A}(\beta)\,d\beta = p_1\int_0^A \beta^{p_1-1}\lambda_f(\beta)\,d\beta.$$

Likewise,

$$\text{(6.30)}\quad \int |Tf|^q\,d\nu = q\int_0^\infty \alpha^{q-1}\lambda_{Tf}(\alpha)\,d\alpha = 2^q q\int_0^\infty \alpha^{q-1}\lambda_{Tf}(2\alpha)\,d\alpha.$$

Since T is sublinear, by Proposition (6.22d) we have

$$\text{(6.31)}\qquad \lambda_{Tf}(2\alpha) \le \lambda_{Tg_A}(\alpha) + \lambda_{Th_A}(\alpha).$$

This is true for all $\alpha > 0$ and all $A > 0$, so we may take A to depend on α. We now make a specific choice of A: by the equations defining p and q, we have

(6.32)

$$\sigma = \frac{p_0(q_0-q)}{q_0(p_0-p)} = \frac{p^{-1}(q^{-1}-q_0^{-1})}{q^{-1}(p^{-1}-p_0^{-1})} = \frac{p^{-1}(q^{-1}-q_1^{-1})}{q^{-1}(p^{-1}-p_1^{-1})} = \frac{p_1(q_1-q)}{q_1(p_1-p)}$$

(the first equation being the definition of σ), and we take $A = \alpha^\sigma$. Then by (6.29), (6.30), (6.31), and the weak type estimates on T,

$$\text{(6.33)}\quad \|Tf\|_q^q \le 2^q q\int_0^\infty \alpha^{q-1}\left[\left(C_0\|g_A\|_{p_0}/\alpha\right)^{q_0} + \left(C_1\|h_A\|_{p_1}/\alpha\right)^{q_1}\right]d\alpha$$

$$\le 2^q q C_0^{q_0} p_0^{q_0/p_0}\int_0^\infty \alpha^{q-q_0-1}\left[\int_{\alpha^\sigma}^\infty \beta^{p_0-1}\lambda_f(\beta)\,d\beta\right]^{q_0/p_0} d\alpha$$

$$+ 2^q q C_1^{q_1} p_1^{q_1/p_1}\int_0^\infty \alpha^{q-q_1-1}\left[\int_0^{\alpha^\sigma} \beta^{p_1-1}\lambda_f(\beta)\,d\beta\right]^{q_1/p_1} d\alpha$$

$$= \sum_{j=0}^{1} 2^q q C_j^{p_j} p_j^{q_j/p_j}\int_0^\infty \left[\int_0^\infty \varphi_j(\alpha,\beta)\,d\beta\right]^{q_j/p_j} d\alpha,$$

where, denoting by χ_0 and χ_1 the characteristic functions of $\{(\alpha,\beta)\colon \beta > \alpha^\sigma\}$ and $\{(\alpha,\beta)\colon \beta < \alpha^\sigma\}$,

$$\varphi_j(\alpha,\beta) = \chi_j(\alpha,\beta)\alpha^{(q-q_j-1)p_j/q_j}\beta^{p_j-1}\lambda_f(\beta).$$

Since $q_0/p_0 \ge 1$ and $q_1/p_1 \ge 1$ we may apply Minkowski's inequality for integrals to obtain

$$\int_0^\infty \left[\int_0^\infty \varphi_j(\alpha,\beta)\,d\beta\right]^{q_j/p_j} d\alpha \le \left[\int_0^\infty \left(\int_0^\infty \varphi_j(\alpha,\beta)^{q_j/p_j}\,d\alpha\right)^{p_j/q_j} d\beta\right]^{q_j/p_j}.$$

Let $\tau = 1/\sigma$. If $q_1 > q_0$, then $q - q_0 > 0$, $\sigma > 0$, and the inequality $\beta > \alpha^\sigma$ is equivalent to $\alpha < \beta^\tau$, so

$$\int_0^\infty \left(\int_0^\infty \varphi_0(\alpha,\beta)^{q_0/p_0}\,d\alpha\right)^{q_0/p_0} d\beta = \int_0^\infty \left(\int_0^{\beta^\tau} \alpha^{q-q_0-1}\,d\alpha\right)^{p_0/q_0} \beta^{p_0-1}\lambda_f(\beta)\,d\beta$$

$$= (q-q_0)^{-p_0/q_0}\int_0^\infty \beta^{p_0-1+p_0(q-q_0)/q_0\sigma}\lambda_f(\beta)\,d\beta$$

$$= (q-q_0)^{-p_0/q_0}\int_0^\infty \beta^{p-1}\lambda_f(\beta)\,d\beta$$

$$= |q-q_0|^{-p_0/q_0}p^{-1}\|f\|_p^p,$$

where we have used (6.32) in simplifying the exponent of β. On the other hand, if $q_1 < q_0$, then $q - q_0 < 0$, $\sigma < 0$, and the inequality $\beta > \alpha^\sigma$ is equivalent to $\alpha > \beta^\tau$, so

$$\int_0^\infty \left(\int_0^\infty \varphi_0(\alpha,\beta)^{q_0/p_0}\,d\alpha\right)^{p_0/q_0} d\beta = \int_0^\infty \left(\int_{\beta^\tau}^\infty \alpha^{q-q_0-1}\,d\alpha\right)^{p_0/q_0} \beta^{p_0-1}\lambda_f(\beta)\,d\beta$$

$$= (q_0-q)^{-p_0/q_0}\int_0^\infty \beta^{p-1}\lambda_f(\beta)\,d\beta$$

$$= |q-q_0|^{-p_0/q_0}p^{-1}\|f\|_p^p.$$

A similar calculation shows that

$$\int_0^\infty \left(\int_0^\infty \varphi_1(\alpha,\beta)^{q_1/p_1}\,d\alpha\right)^{p_1/q_1} d\beta = |q-q_1|^{-p_1/q_1}p^{-1}\|f\|_p^p.$$

Combining these results with (6.33), we see that

$$\sup\{\|Tf\|_q\colon \|f\|_p = 1\} \le B_p,$$

$$B_p = 2q^{1/q}\left[\sum_{j=0}^{1} C_j^{q_j}\left(p_j/p\right)^{q_j/p_j}|q-q_j|^{-1}\right]^{1/q}.$$

But since $|T(cf)| = |c||Tf|$, this implies that $\|Tf\|_q \le B_p\|f\|_p$ for all $f \in L^p(\mu)$, and we are done. (The verification of the asserted properties of B_p is left as an easy exercise.)

It remains to show how to modify this argument to deal with the exceptional cases $q_0 = \infty$ or $q_1 = \infty$. We distinguish three cases.

Case I: $p_1 = q_1 = \infty$ (so $p_0 \le q_0 < \infty$). Instead of taking $A = \alpha^\sigma$ in the decomposition of f, we take $A = \alpha/C_1$. Then $\|Th_A\|_\infty \le C_1\|h_A\|_\infty \le \alpha$, so $\lambda_{Th_A}(\alpha) = 0$, and we obtain (6.33) with $\varphi_1 = 0$ and α^σ replaced by α/C_1 in the definition of φ_0. The same argument as above then gives

$$\|Tf\|_q \le 2\Big[qC_0^{q_0}C_1^{q-q_0}(p_0/p)^{q_0/p_0}|q-q_0|^{-1}\Big]^{1/q}\|f\|_p.$$

Case II: $p_0 < p_1 < \infty$, $q_0 < q_1 = \infty$. Again the idea is to choose A so that $\lambda_{Th_A}(\alpha) = 0$, and the proper choice is $A = (\alpha/d)^\sigma$ where

$$d = C_1\Big[p_1\|f\|_p^p/p\Big]^{1/p_1}, \quad \sigma = p_1/(p_1 - p).$$

[This value of σ is the limiting case of (6.32) as $q_1 \to \infty$.] Indeed, since $p_1 < p$ we have

$$\|Th_A\|_\infty^{p_1} \le C_1^{p_1}\|h_A\|_{p_1}^{p_1} = C_1^{p_1}p_1\int_0^A \alpha^{p_1-1}\lambda_f(\alpha)\,d\alpha$$

$$\le C_1^{p_1}p_1A^{p_1-p}\int_0^A \alpha^{p-1}\lambda_f(\alpha)\,d\alpha \le C_1^{p_1}(p_1/p)(\alpha/d)^{p_1}\|f\|_p^p = \alpha^{p_1}.$$

As in Case I, then, we find that $\varphi_1 = 0$ in (6.33) and the integral involving φ_0 is majorized by a constant B_p when $\|f\|_p = 1$, which yields the desired result.

Case III: $p_0 < p_1 < \infty$, $q_1 < q_0 = \infty$. The argument is essentially the same as in Case II, except that we take $A = (\alpha/d)^\sigma$ with d chosen so that $\lambda_{Tg_A}(\alpha) = 0$. □

For the reader who is overwhelmed by this proof, we recommend the exercise of working out two special cases: (i) $p_0 = q_0 = 1$, $p_1 = q_1 = 2$, and (ii) $p_0 = q_0 = 1$, $p_1 = q_1 = \infty$.

Let us compare our two interpolation theorems. The Marcinkiewicz theorem requires some restrictions on p_j and q_j which are not present in the Riesz–Thorin theorem; these restrictions, however, are satisfied in all the interesting applications. Apart from this, the hypotheses of the Marcinkiewicz theorem are weaker: T is allowed to be sublinear rather than linear, and it needs only to satisfy weak-type estimates at the endpoints. The conclusion in both cases is that T is bounded from $L^p(\mu)$ to $L^q(\nu)$, but the Riesz–Thorin theorem produces a much sharper estimate for the operator norm of T. Thus neither theorem includes the other.

We conclude with two applications of the Marcinkiewicz theorem. The first concerns the Hardy–Littlewood maximal operator H discussed in Section 3.4. H is obviously sublinear and satisfies $\|Hf\|_\infty \le \|f\|_\infty$ for all $f \in L^\infty$. Moreover, Theorem (3.17) says precisely that H is weak type $(1,1)$. We conclude:

(6.34) Corollary. There is a constant $C > 0$ such that if $1 < p < \infty$ and $f \in L^p(\mathbf{R}^n)$, then

$$\|Hf\|_p \le Cp(p-1)^{-1}\|f\|_p. \qquad \square$$

Our second application is a theorem on integral operators related to Theorem (6.18).

(6.35) Theorem. Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces, and $1 < q < \infty$. Let K be a measurable function on $X \times Y$ such that, for some $C > 0$, we have $[K(x,\cdot)]_q \le C$ for a.e. $x \in X$ and $[K(\cdot,y)]_q \le C$ for a.e. $y \in Y$. If $1 \le p < \infty$ and $f \in L^p(\nu)$, the integral

$$Tf(x) = \int K(x,y)f(y)\,d\nu(y)$$

converges absolutely for a.e. $x \in X$, and the operator T thus defined is weak type $(1,q)$ and strong type (p,r) provided $1 < p < r < \infty$ and $p^{-1} + q^{-1} = r^{-1} + 1$. More precisely, there exist constants A, B_p independent of K such that $[Tf]_q \le AC\|f\|_1$ and $\|Tf\|_r \le B_pC\|f\|_p$ for $p > 1$.

Proof. Let p', q' be the conjugate exponents to p, q; then the equation $p^{-1} + q^{-1} = r^{-1} + 1$ can be restated as $r^{-1} = p^{-1} - (q')^{-1} = q^{-1} - (p')^{-1}$, so that for $r < \infty$ we have $p < q'$ and $q < p'$. If $0 \ne f \in L^p$ $(1 \le p < q')$, by multiplying f and K by constants we may assume that $\|f\|_p = C = 1$. Given a positive number A whose value will be fixed later, set $E = \{(x,y): |K(x,y)| > A\}$, define $K_1 = (\operatorname{sgn} K)(|K| - A)\chi_E$ and $K_2 = K - K_1$, and let T_1, T_2 be the operators corresponding to K_1, K_2. Then by Propositions (6.24) and (6.25), since $q > 1$ we have

$$\int|K_1(x,y)|\,d\nu(y) = \int_0^\infty \lambda_{K(x,\cdot)}(\alpha + A)\,d\alpha \le \int_A^\infty \alpha^{-q}\,d\alpha = \frac{A^{1-q}}{q-1},$$

and likewise

$$\int|K_1(x,y)|\,d\mu(x) \le \frac{A^{1-q}}{q-1}.$$

Hence by Theorem (6.18), the integral defining $T_1f(x)$ converges for a.e. x and

$$\|T_1f\|_p \le \frac{A^{1-q}}{q-1}\|f\|_p = \frac{A^{1-q}}{q-1}. \tag{6.36}$$

Similarly, since $q < p'$,

$$\int |K_2(x,y)|^{p'}\,d\nu(y) = p'\int_0^A \alpha^{p'-1}\lambda_{K(x,\cdot)}(\alpha)\,d\alpha$$

$$\leq p'\int_0^A \alpha^{p'-1-q}\,d\alpha = \frac{p'}{p'-q}A^{p'-q}.$$

Therefore, by Hölder's inequality, the integral defining $T_2 f(x)$ converges for every x, and

$$\|T_2 f\|_\infty \leq \left(\frac{p'}{p'-q}\right)^{1/p'} A^{(p'-q)/p'}\|f\|_p = \left(\frac{r}{q}\right)^{1/p'} A^{q/r}. \tag{6.37}$$

We have thus established that $Tf = T_1 f + T_2 f$ is well defined. Given $\alpha > 0$, we wish to estimate $\lambda_{Tf}(\alpha)$. But by Proposition (6.22d),

$$\lambda_{Tf}(\alpha) \leq \lambda_{T_1 f}(\alpha/2) + \lambda_{T_2 f}(\alpha/2),$$

and by (6.37), if we choose

$$A = \left(\frac{\alpha}{2}\right)^{r/q}\left(\frac{q}{r}\right)^{r/qp'}$$

we will have $\|T_2 f\|_\infty \leq \alpha/2$, so that $\lambda_{T_2 f}(\alpha/2) = 0$. With this choice of A, then, by (6.36) and Chebyshev's inequality we obtain

$$\lambda_{Tf}(\alpha) \leq \lambda_{T_1 f}\left(\frac{\alpha}{2}\right) \leq \left(\frac{2\|T_1 f\|_p}{\alpha}\right)^p \leq \left(\frac{2}{q-1}\right)^p A^{(1-q)p}\alpha^{-p}$$

$$= \left(\frac{2}{q-1}\right)^p \left(\frac{q}{r}\right)^{r(1-q)p/qp'} \left(\frac{\alpha}{2}\right)^{(1-q)pr/q} \alpha^{-p}$$

$$= C_p\left(\frac{\|f\|_p}{\alpha}\right)^r$$

because $\|f\|_p = 1$ and

$$\frac{(1-q)pr}{q} - p = p\left(-\frac{r}{q'} - 1\right) = p\left(\frac{-r}{p}\right) = -r.$$

A simple homogeneity argument now shows that the estimate $\lambda_{Tf}(\alpha) \leq C_p(\|f\|_p/\alpha)^r$ holds with no restriction on $\|f\|_p$, so we have shown that T is weak type (p,r), and in particular (for $p = 1$) weak type $(1,q)$. Finally, given

p with $1 < p < q'$, choose $\hat{p}$ such that $p < \hat{p} < q'$ and define $\hat{r}$ by $\hat{r}^{-1} = \hat{p}^{-1} - (q')^{-1}$. Then T is weak types $(1,q)$ and $(\hat{p},\hat{r})$, so it follows from the Marcinkiewicz theorem that T is strong type (p,r). □

EXERCISES

41. Suppose $1 < p \leq \infty$ and $p^{-1} + q^{-1} = 1$. If T is a bounded operator on L^p such that $\int (Tf)g = \int f(Tg)$ for all $f, g \in L^p \cap L^q$, then T extends uniquely to a bounded operator on L^r for all r such that $p \leq r \leq q$ (if $p < q$) or $q \leq r \leq p$ (if $p > q$).

42. Prove the Marcinkiewicz theorem in the case $p_0 = p_1$. [Setting $p = p_0 = p_1$, we have $\lambda_{Tf}(\alpha) \leq (C_0\|f\|_p/\alpha)^{q_0}$ and $\lambda_{Tf}(\alpha) \leq (C_1\|f\|_p/\alpha)^{q_1}$. Use whichever estimate is better, depending on α, to majorize $q\int_0^\infty \alpha^{q-1}\lambda_{Tf}(\alpha)\,d\alpha$.]

43. Let H be the Hardy–Littlewood maximal operator on $\mathbf{R}$. Compute $H\chi_{(0,1)}$ explicitly. Show that it is in L^p for $p > 1$ and in weak L^1 but not in L^1, and that its L^p norm tends to ∞ like $(p-1)^{-1}$ as $p \to 1$, although $\|\chi_{(0,1)}\|_p = 1$ for all p.

44. Let I_α be the fractional integration operator of Exercise 2.63. If $0 < \alpha < 1$, then I_α is weak type $(1,(1-\alpha)^{-1})$ and strong type (p,r) with respect to Lebesgue measure on $(0,\infty)$ where $1 < p < \alpha^{-1}$ and $r^{-1} = p^{-1} - \alpha$.

45. If $0 < \alpha < n$, define an operator T_α on functions on $\mathbf{R}^n$ by

$$T_\alpha f(x) = \int |x-y|^{-\alpha} f(y)\,dy.$$

Then T_α is weak type $(1, n(n-\alpha)^{-1})$ and strong type (p,r) with respect to Lebesgue measure on $\mathbf{R}^n$ where $1 < p < n\alpha^{-1}$ and $r^{-1} = p^{-1} - \alpha n^{-1}$. [The case $n = 3$, $\alpha = 1$ is of particular interest in physics: if f represents the density of a mass or charge distribution, $-(4\pi)^{-1}T_1 f$ represents the induced gravitational or electrostatic potential.]

6. NOTES AND REFERENCES

The importance of the space $L^2([a,b])$ was recognized soon after the invention of the Lebesgue integral because of its connection with Fourier series and other orthogonal expansions; and one of the early triumphs of the Lebesgue theory was the discovery in 1907 by Fischer [31] and F. Riesz [80] that $L^2([a,b])$ is isomorphic to l^2, or, what amounts to the same thing, that $L^2([a,b])$ is complete. The spaces $L^p([a,b])$ for $1 < p < \infty$ were first investigated by F. Riesz [83], who proved all of the major results in Sections 1 and 2 for them as

well as the weak sequential compactness of the closed unit ball in L^p. The fact that $(L^1)^* \cong L^\infty$ was first proved by Steinhaus [102].

In retrospect it seems somewhat unfortunate that L^p spaces were not named $L^{1/p}$ spaces, for—as one sees in the conjugacy relation $p^{-1} + q^{-1} = 1$ and in the results of Section 5—relationships among different L^p spaces usually involve linear equations in p^{-1}.

Section 1. Hölder's inequality, in the case $p = 2$, is commonly associated with the names of Cauchy (who proved it for finite sums), and Buniakovsky and Schwarz (who proved it, independently, for integrals). For general p it was discovered independently by Hölder and Rogers. (See Hardy, Littlewood, and Pólya [45] for references.) A neat proof of Hölder's inequality using complex function theory can be found in Rubel [88].

For a discussion of Exercise 5, including some other conditions for the inclusion $L^p \subset L^q$ to hold, see Romero [86].

Section 2. A quite different approach to the L^p duality theory for $1 < p < \infty$ can be found in Hewitt and Stromberg [54], Section 15. J. Schwartz [94] has found a characterization of $L^1(\mu)^*$ which is valid for arbitrary measures μ.

The proof of Theorem (6.15) breaks down for $p = \infty$ because the set function $\nu(E) = \varphi(\chi_E)$ need not be countably additive. It is, however, a bounded, *finitely* additive complex measure on $(X, \mathcal{M})$ which is absolutely continuous with respect to μ in the sense that $\nu(E) = 0$ whenever $\mu(E) = 0$. Conversely, given a bounded, finitely additive complex measure ν on $(X, \mathcal{M})$, one can define the integral of a bounded measurable function with respect to ν. (One defines $\int f d\nu$ when f is simple in the obvious way and then shows that $|\int f d\nu| \le C\|f\|_u$, so that the integral extends to all uniform limits of simple functions.) In this way one obtains a representation of $(L^\infty)^*$ as a space of finitely additive complex measures. See Hewitt and Stromberg [54], Section 20, and, for a more general treatment of finitely additive integrals, Dunford and Schwartz [26], Chapter 3. [The example of a $\varphi \in (L^\infty)^* \setminus L^1$ which we presented at the end of Section 2 shows how horrible finitely additive measures can be: if $\nu(E) = \varphi(\chi_E)$, then $\nu \ll m$, but ν behaves like the point mass at zero when integrated against any continuous function.]

Section 3. Theorem (6.18) appears to be a folk theorem; it generalizes a classical result of W. H. Young which we present in Section 8.2. The reader whose appetite for inequalities is not satisfied by this section can find a feast in Hardy, Littlewood, and Pólya [45].

Section 5. The Riesz–Thorin theorem was first proved by M. Riesz [84] (F. Riesz's younger brother) under the assumption that $p_j \le q_j$ $(j = 0, 1)$; the proof in the general case and the idea of using the three lines lemma are due to Thorin [107]. E. M. Stein has proved a very powerful generalization of the Riesz–Thorin theorem. It deals with a family $\{T_z: 0 \le \operatorname{Re} z \le 1\}$ of operators

which (roughly speaking) depend holomorphically on z and satisfy some mild growth conditions as $|\operatorname{Im} z| \to \infty$, and it asserts that if T_z is bounded from L^{p_j} to L^{q_j} for $\operatorname{Re} z = j$ $(j = 0, 1)$, then T_z is bounded from L^{p_t} to L^{q_t} for $\operatorname{Re} z = t$ $(0 < t < 1)$, where p_t, q_t are defined as in the Riesz–Thorin theorem. The precise statement and proof may be found in Stein and Weiss [101], Section V.4, or Zygmund [113], Section XII.1; for a further extension of these ideas, see Coifman et al. [20].

The Marcinkiewicz interpolation theorem was announced by Marcinkiewicz [68] for the case $p_j = q_j$ $(j = 0, 1)$; after his untimely death in World War II, the work was completed by Zygmund [112]. The theorem can be proved under still weaker hypotheses on T: an extra twist to the argument we have given yields the result under the sole assumption that $|T(f + g)| \le C(|Tf| + |Tg|)$ for some constant C. See Zygmund [112] and [113], Section XII.4. The spaces L^p and weak L^p form part of a two-parameter family $\{L(p, q): 1 \le p, q \le \infty\}$ of function spaces, the so-called Lorentz spaces, such that $L^p = L(p, p)$ and weak $L^p = L(p, \infty)$, and the Marcinkiewicz theorem can be extended to a result about interpolation of operators on the $L(p, q)$ spaces. See Stein and Weiss [101], Section V.3.

There are many other examples of "continuous families" of Banach spaces for which interpolation theorems can be proved—for example, the spaces Λ_α discussed in Exercise 5.9 and the Sobolev spaces discussed in Section 8.6. There are also two general techniques for constructing "intermediate spaces" between pairs of Banach spaces, known as the "complex method" and the "real method," which may be regarded as abstract forms of the Riesz–Thorin and Marcinkiewicz theorems. An account of these matters can be found in Bergh and Löfström [10].

Theorem (6.35) appears first in Folland and Stein [34], but the essential idea of the proof was discovered by Stein several years earlier (see, for example, Stein [100], Section V.1), and the special case discussed in Exercise 44 goes back to Hardy and Littlewood [43].

7

RADON MEASURES

The subject of this chapter is measure and integration theory on locally compact Hausdorff (LCH) spaces. We have seen in Section 2.6 that Lebesgue measure on $\mathbf{R}^n$ interacts nicely with the topology on $\mathbf{R}^n$—measurable sets can be approximated by open or compact sets, and integrable functions can be approximated by continuous functions—and it is of interest to study measures having similar properties on more general spaces. Moreover, it turns out that certain linear functionals on spaces of continuous functions are given by integration against such measures. This fact constitutes an important link between measure theory and functional analysis, and also provides a powerful tool for constructing measures.

Throughout this chapter, X will denote an LCH space. We continue to employ the terminology developed in Chapter 1 in the context of metric spaces: $\mathscr{B}_X$ will denote the Borel σ-algebra on X, that is, the σ-algebra generated by the open sets; measures on $\mathscr{B}_X$ will be called Borel measures; countable unions (intersections) of closed (open) sets will be called F_σ (G_δ) sets, and so forth.

1. POSITIVE LINEAR FUNCTIONS ON $C_c(X)$

A linear functional I on $C_c(X)$ will be called **positive** if $I(f) \geq 0$ whenever $f \geq 0$. In this definition there is no mention of continuity, but it is worth noting that positivity itself implies a rather strong continuity property.

(7.1) Proposition. If I is a positive linear functional on $C_c(X)$, for each compact $K \subset X$ there is a constant C_K such that $|I(f)| \leq C_K \|f\|_u$ for all $f \in C_c(X)$ such that $\operatorname{supp}(f) \subset K$.

Proof. It suffices to consider real-valued f. Given a compact K, choose $\varphi \in C_c(X, [0,1])$ such that $\varphi = 1$ on K (Urysohn's lemma). Then if $\operatorname{supp}(f) \subset$

K, we have $|f| \leq \|f\|_u \varphi$, that is, $\|f\|_u \varphi - f \geq 0$ and $\|f\|_u \varphi + f \geq 0$. Thus $\|f\|_u I(\varphi) - I(f) \geq 0$ and $\|f\|_u I(\varphi) + I(f) \geq 0$, so that $|I(f)| \leq I(\varphi)\|f\|_u$. □

If μ is a Borel measure on X such that $\mu(K) < \infty$ for every compact $K \subset X$, then clearly $C_c(X) \subset L^1(\mu)$, so the map $f \to \int f d\mu$ is a positive linear functional on $C_c(X)$. The principal result of this section is that *every* positive linear functional on $C_c(X)$ arises in this fashion; moreover, one can impose some additional regularity conditions on μ, subject to which μ is unique. These conditions are as follows.

Let μ be a Borel measure on X and E a Borel subset of X. μ is called **outer regular** on E if

$$\mu(E) = \inf\{\mu(U) : U \supset E, U \text{ open}\}$$

and **inner regular** on E if

$$\mu(E) = \sup\{\mu(K) : K \subset E, K \text{ compact}\}.$$

If μ is outer and inner regular on all Borel sets, μ is called **regular**. It turns out that regularity is a bit too much to ask for when X is not σ-compact, so we adopt the following definition. A **Radon measure** on X is a Borel measure which is finite on compact sets, outer regular on all Borel sets, and inner regular on all open sets. We shall show in the next section that Radon measures are also inner regular on all of their σ-finite sets.

One further bit of notation: if U is open in X and $f \in C_c(X)$, we say that $f \prec U$ (read: f is **subordinate** to U) if $0 \leq f \leq 1$ and $\operatorname{supp}(f) \subset U$. [This is slightly stronger than the condition $0 \leq f \leq \chi_U$, which means that $\operatorname{supp}(f) \subset \overline{U}$.]

(7.2) The Riesz Representation Theorem. If I is a positive linear functional on $C_c(X)$, there is a unique Radon measure μ on X such that $I(f) = \int f d\mu$ for all $f \in C_c(X)$. Moreover, μ satisfies

(7.3) $\mu(U) = \sup\{I(f) : f \in C_c(X) \text{ and } f \prec U\}$ for all open $U \subset X$,

(7.4) $\mu(K) = \inf\{I(f) : f \in C_c(X) \text{ and } f \geq \chi_K\}$ for all compact $K \subset X$.

Proof. Let us begin by establishing uniqueness. If μ is a Radon measure such that $I(f) = \int f d\mu$ for all $f \in C_c(X)$, and $U \subset X$ is open, then clearly $I(f) \leq \mu(U)$ whenever $f \prec U$. On the other hand, if $K \subset U$ is compact, by Urysohn's lemma there is an $f \in C_c(X)$ such that $f \prec U$ and $f = 1$ on K, whence $\mu(K) \leq \int f d\mu = I(f)$. Since μ is inner regular on U, it follows that (7.3) is satisfied. Thus μ is determined by I on open sets, and hence on all Borel sets because of outer regularity.

This argument proves the uniqueness of μ and also suggest how to go about proving existence. Namely, we begin by defining

$$\mu(U) = \sup\{I(f): f \in C_c(X) \text{ and } f \prec U\}$$

for U open, and we then define $\mu^*(E)$ for an arbitrary $E \subset X$ by

$$\mu^*(E) = \inf\{\mu(U): U \supset E, U \text{ open}\}.$$

Clearly $\mu(U) \leq \mu(V)$ if $U \subset V$, and hence $\mu(U) = \mu^*(U)$ when U is open. The outline of the proof is now as follows. We shall establish that:

(I) μ^* is an outer measure.

(II) Every open set is μ^*-measurable.

At this point it follows from Carathéodory's theorem that every Borel set is μ^*-measurable and that $\mu = \mu^*|\mathscr{B}_X$ is a Borel measure. μ is outer regular and satisfies (7.3) by definition. We next show:

(III) μ satisfies (7.4).

This clearly implies that μ is finite on compact sets, and inner regularity on open sets also follows easily. Indeed, if U is open and $\alpha < \mu(U)$, choose $f \in C_c(X)$ such that $f \prec U$ and $I(f) > \alpha$, and let $K = \text{supp}(f)$. If $g \in C_c(X)$ and $g \geq \chi_K$, then $g \geq f$ and hence $I(g) \geq I(f) > \alpha$. But then $\mu(K) > \alpha$ by (7.4), so μ is inner regular on U. Finally, we prove:

(IV) $I(f) = \int f d\mu$ for all $f \in C_c(X)$.

With this, the proof of the theorem will be complete.

Proof of (I): It suffices to show that if $\{U_j\}$ is a sequence of open sets and $U = \bigcup_1^\infty U_j$, then $\mu(U) \leq \sum_1^\infty \mu(U_j)$. Indeed, from this it follows that for any $E \subset X$,

$$\mu^*(E) = \inf\left\{\sum_1^\infty \mu(U_j): U_j \text{ open}, E \subset \bigcup_1^\infty U_j\right\},$$

and the expression on the right defines an outer measure by Proposition (1.10). If $U = \bigcup_1^\infty U_j$, $f \in C_c(X)$, and $f \prec U$, let $K = \text{supp}(f)$. Since K is compact we have $K \subset \bigcup_1^n U_j$ for some finite n, so by Proposition (4.41) there exist $g_1, \ldots, g_n \in C_c(X)$ with $g_j \prec U_j$ and $\sum_1^n g_j = 1$ on K. But then $f = \sum_1^n f g_j$ and $f g_j \prec U_j$, so

$$I(f) = \sum_1^n I(f g_j) \leq \sum_1^n \mu(U_j) \leq \sum_1^\infty \mu(U_j).$$

Since this is true for any $f \prec U$, we conclude that $\mu(U) \leq \sum_1^\infty \mu(U_j)$ as desired.

Proof of (II): We must show that if U is open and E is any subset of X such that $\mu^*(E) < \infty$, then $\mu^*(E) \geq \mu^*(E \cap U) + \mu^*(E \setminus U)$. First suppose that E is open. Then $E \cap U$ is open, so given $\varepsilon > 0$ we can find $f \in C_c(X)$ such that $f \prec E \cap U$ and $I(f) > \mu(E \cap U) - \varepsilon$. Also, $E \setminus (\text{supp}(f))$ is open, so we can find $g \in C_c(X)$ such that $g \prec E \setminus \text{supp}(f)$ and $I(g) > \mu(E \setminus \text{supp}(f)) - \varepsilon$. But then $f + g \prec E$, so

$$\begin{aligned}\mu(E) &\geq I(f) + I(g) > \mu(E \cap U) + \mu(E \setminus \text{supp}(f)) - 2\varepsilon \\ &\geq \mu^*(E \cap U) + \mu^*(E \setminus U) - 2\varepsilon.\end{aligned}$$

Letting $\varepsilon \to 0$ we obtain the desired inequality. For the general case, if $\mu^*(E) < \infty$ we can find an open $V \supset E$ such that $\mu(V) < \mu^*(E) + \varepsilon$, and hence

$$\begin{aligned}\mu^*(E) + \varepsilon > \mu(V) &\geq \mu^*(V \cap U) + \mu^*(V \setminus U) \\ &\geq \mu^*(E \cap U) + \mu^*(E \setminus U).\end{aligned}$$

Letting $\varepsilon \to 0$, we are done.

Proof of (III). If K is compact, $f \in C_c(X)$, and $f \geq \chi_K$, let $U_\varepsilon = \{x: f(x) > 1 - \varepsilon\}$. Then U_ε is open, and if $g \prec U_\varepsilon$, we have $g \leq (1 - \varepsilon)^{-1} f$ and so $I(g) \leq (1 - \varepsilon)^{-1} I(f)$. Thus $\mu(K) \leq \mu(U_\varepsilon) \leq (1 - \varepsilon)^{-1} I(f)$, and letting $\varepsilon \to 0$ we see that $\mu(K) \leq I(f)$. On the other hand, for any open $U \supset K$, by Urysohn's lemma there exists $f \in C_c(X)$ such that $f \geq \chi_K$ and $f \prec U$, whence $I(f) \leq \mu(U)$. Since μ is outer regular on K, (7.4) follows.

Proof of (IV). It suffices to show that $I(f) = \int f d\mu$ if $f \in C_c(X, [0,1])$, since $C_c(X)$ is the linear span of the latter set. Given $N \in \mathbb{N}$, for $1 \leq j \leq N$ let $K_j = \{x: f(x) \geq N^{-1} j\}$ and let $K_0 = \text{supp}(f)$. Also, define $f_1, \ldots, f_N \in C_c(X)$ by $f_i(x) = 0$ if $x \notin K_{j-1}$, $f_j(x) = f(x) - N^{-1}(j-1)$ if $x \in K_{j-1} \setminus K_j$, and $f_j(x) = N^{-1}$ if $x \in K_j$. In other words,

$$f_j = \min\{\max\{f - N^{-1}(j-1), 0\}, N^{-1}\}.$$

Then $N^{-1}\chi_{K_j} \leq f_j \leq N^{-1}\chi_{K_{j-1}}$, hence

$$N^{-1}\mu(K_j) \leq \int f_j d\mu \leq N^{-1}\mu(K_{j-1}),$$

Also, $f_j \in C_c(X)$, so if U is an open set containing K_{j-1} we have $I(f_j) \leq N^{-1}\mu(U)$ and hence, by (7.4) and outer regularity,

$$N^{-1}\mu(K_j) \leq I(f_j) \leq N^{-1}\mu(K_{j-1}).$$

Moreover, $f = \sum_1^N f_j$, so that

$$N^{-1}\sum_1^N \mu(K_j) \le \int f\,d\mu \le N^{-1}\sum_0^{N-1} \mu(K_j),$$

$$N^{-1}\sum_1^N \mu(K_j) \le I(f) \le N^{-1}\sum_0^{N-1} \mu(K_j).$$

It follows that

$$|I(f) - \int f\,d\mu| \le N^{-1}[\mu(K_0) - \mu(K_N)] \le N^{-1}\mu(\operatorname{supp}(f)).$$

Since $\mu(\operatorname{supp}(f)) < \infty$ and N is arbitrary we conclude that $I(f) = \int f\,d\mu$. □

The proof of this theorem yields something stronger than the statement: we obtain not just a Borel measure μ but an extension $\bar{\mu}$ of μ to the σ-algebra of μ^*-measurable sets. However, it follows from outer regularity that for any $E \subset X$,

$$\mu^*(E) = \inf\{\mu(B)\colon B \in \mathscr{B}_X, B \supset E\},$$

so μ^* is the outer measure induced by μ in the sense of Section 1.4. According to Exercise 1.22, $\bar{\mu}$ is therefore the completion of μ if μ is σ-finite and is the saturation of the completion of μ in general.

On the other hand, some authors prefer to restrict attention to a smaller σ-algebra than $\mathscr{B}_X$, namely the σ-algebra $\mathscr{B}_X^0$ generated by $C_c(X)$ [that is, the smallest σ-algebra with respect to which every $f \in C_c(X)$ is measurable]. The elements of $\mathscr{B}_X^0$ are called **Baire sets**. For more about Baire sets, see Exercises 4–6.

EXERCISES

1. Let X be an LCH space, Y a closed subset of X (which is an LCH space in the relative topology), and μ a Radon measure on Y. Then $I(f) = \int (f|Y)\,d\mu$ is a positive linear functional on $C_c(X)$, and its induced Radon measure ν on X is given by $\nu(E) = \mu(E \cap Y)$.
2. Let μ be a Radon measure on X.
   (a) Let N be the union of all open $U \subset X$ such that $\mu(U) = 0$. Then N is open, $\mu(N) = 0$, and if V is open and $V \setminus N \neq \varnothing$ then $\mu(V) > 0$. The complement of N is called the **support** of μ.
   (b) $x \in \operatorname{supp}(\mu)$ iff $\int f\,d\mu > 0$ for every $f \in C_c(X, [0,1])$ such that $f(x) > 0$.

3. Let X be the one-point compactification of a set with the discrete topology. If μ is a Radon measure on X, then $\operatorname{supp}(\mu)$ (see Exercise 2) is countable.
4. Let X be an LCH space.
   (a) If $f \in C_c(X, [0,\infty))$, then $f^{-1}([a,\infty))$ is a compact G_δ set for all $a > 0$.
   (b) If $K \subset X$ is a compact G_δ set, there exists $f \in C_c(X, [0,1])$ such that $K = f^{-1}(\{1\})$.
   (c) The σ-algebra $\mathscr{B}_X^0$ of Baire sets is the σ-algebra generated by the compact G_δ sets.
5. Let X be a second countable LCH space.
   (a) Every compact subset of X is a G_δ set.
   (b) $\mathscr{B}_X = \mathscr{B}_X^0$.
6. Let X be an uncountable set with the discrete topology, or the one-point compactification of such a set. Then $\mathscr{B}_X \neq \mathscr{B}_X^0$.

2. REGULARITY AND APPROXIMATION THEOREMS

In this section we explore in more detail the properties of Radon measures.

(7.5) Proposition. If μ is a Radon measure on X, μ is inner regular on all of its σ-finite sets.

Proof. Suppose $\mu(E) < \infty$. Given $\varepsilon > 0$, we can choose an open $U \supset E$ such that $\mu(U) < \mu(E) + \varepsilon$ and a compact $F \subset U$ such that $\mu(F) > \mu(U) - \varepsilon$. Since $\mu(U \setminus E) < \varepsilon$ we can also choose an open $V \supset U \setminus E$ such that $\mu(V) < \varepsilon$. Let $K = F \setminus V$. Then K is compact, $K \subset E$, and

$$\mu(K) = \mu(F) - \mu(F \cap V) > \mu(E) - \varepsilon - \mu(V) > \mu(E) - 2\varepsilon.$$

Thus μ is inner regular on E. On the other hand, if E is σ-finite but $\mu(E) = \infty$, E is an increasing union of sets E_j with $\mu(E_j) < \infty$ and $\mu(E_j) \to \infty$. Thus for any $N \in \mathbf{N}$ there exists j such that $\mu(E_j) > N$ and hence, by the preceding argument, a compact $K \subset E_j$ with $\mu(K) > N$. Hence μ is inner regular on E. □

(7.6) Corollary. Every σ-finite Radon measure on X is regular. If X is σ-compact, every Radon measure on X is regular.

For an example of a non-regular Radon measure on a non-σ-compact space, see Exercise 13.

(7.7) Proposition. Suppose μ is a σ-finite Radon measure on X, and E is a Borel set in X.

(a) For every $\varepsilon > 0$ there exist an open U and a closed F with $F \subset E \subset U$ and $\mu(U \setminus F) < \varepsilon$.

(b) There exist an F_σ set A and a G_δ set B such that $A \subset E \subset B$ and $\mu(B \setminus A) = 0$.

Proof. Write $E = \bigcup_1^\infty E_j$ where the E_j's are disjoint and have finite measure. For each j, choose an open $U_j \supset E_j$ with $\mu(U_j) < \mu(E_j) + \varepsilon 2^{-j-1}$ and let $U = \bigcup_1^\infty U_j$. Then U is open, $U \supset E$, and $\mu(U \setminus E) \le \sum_1^\infty \mu(U_j \setminus E_j) < \varepsilon/2$. Applying the same reasoning to E^c, we obtain an open $V \supset E^c$ with $\mu(V \setminus E^c) < \varepsilon/2$. Let $F = V^c$. Then F is closed, $F \subset E$, and

$$\mu(U \setminus F) = \mu(U \setminus E) + \mu(E \setminus F) = \mu(U \setminus E) + \mu(V \setminus E^c) < \varepsilon.$$

This proves (a), and (b) follows easily; details are left to the reader. □

(7.8) Theorem. Let X be an LCH space in which every open set is σ-compact (which is the case, for example, if X is second countable). Then every Borel measure on X which is finite on compact sets is regular and hence Radon.

Proof. If μ is a Borel measure which is finite on compact sets, then $C_c(X) \subset L^1(\mu)$, so the map $I(f) = \int f d\mu$ is a positive linear functional on $C_c(X)$. Let ν be the associated Radon measure according to Theorem (7.2). If $U \subset X$ is open, let $U = \bigcup_1^\infty K_j$ where each K_j is compact. Choose $f_1 \in C_c(X)$ so that $f_1 \prec U$ and $f_1 = 1$ on K_1. Proceeding inductively, for $n > 1$ choose $f_n \in C_c(X)$ so that $f_n \prec U$ and $f_n = 1$ on $\bigcup_1^n K_j$ and on $\bigcup_1^{n-1} \operatorname{supp}(f_j)$. Then f_n increases pointwise to χ_U as $n \to \infty$, so

$$\mu(U) = \lim \int f_n \, d\mu = \lim \int f_n \, d\nu = \nu(U)$$

by the monotone convergence theorem. Next, if E is any Borel set and $\varepsilon > 0$, by Proposition (7.7) there exist an open $V \supset E$ and a closed $F \subset E$ with $\nu(V \setminus F) < \varepsilon$. But $V \setminus F$ is open, so $\mu(V \setminus F) < \varepsilon$. In particular, $\mu(V) \le \mu(E) + \varepsilon$, so μ is outer regular. Also, $\mu(F) \ge \mu(E) - \varepsilon$, and F is σ-compact (since X is), so there exist compact $K_j \subset F$ with $\mu(K_j) \to \mu(F)$, whence μ is inner regular. Thus μ is regular [and equal to ν, by the uniqueness part of Theorem (7.2)]. □

For an example of a finite, non-Radon Borel measure on a compact Hausdorff space, see Exercise 15.

We now turn to some approximation theorems for measurable functions.

(7.9) Proposition. If μ is a Radon measure on X, $C_c(X)$ is dense in $L^p(\mu)$ for $1 \le p < \infty$.

Proof. Since the L^p simple functions are dense in L^p [Proposition (6.7)], it suffices to show that for any Borel set E with $\mu(E) < \infty$, χ_E can be approximated in the L^p norm by elements of $C_c(X)$. Given $\varepsilon > 0$, by Proposition (7.5) we can choose a compact $K \subset E$ and an open $U \supset E$ such that $\mu(U \setminus K) < \varepsilon$, and by Urysohn's lemma we can choose an $f \in C_c(X)$ such that $\chi_K \le f \le \chi_U$. Then

$$\|\chi_E - f\|_p \le \mu(U \setminus K)^{1/p} < \varepsilon^{1/p}. \qquad \square$$

(7.10) Lusin's Theorem. Suppose μ is a Radon measure on X and $f: X \to \mathbf{C}$ is a measurable function such that $\mu(E) < \infty$ where $E = \{x: f(x) \neq 0\}$. For any $\varepsilon > 0$ there exists $\varphi \in C_c(X)$ such that $\mu(\{x: f(x) \neq \varphi(x)\}) < \varepsilon$. If f is bounded, φ can be taken to satisfy $\|\varphi\|_u \le \|f\|_u$.

Proof. First suppose f is bounded. Then $f \in L^1(\mu)$, so by Proposition (7.9) there is a sequence $\{g_n\}$ in $C_c(X)$ which converges to f in L^1, and hence by Corollary (2.32) a subsequence (still denoted by $\{g_n\}$) which converges to f a.e. By Egoroff's theorem there is a set $A \subset E$ such that $\mu(A \setminus E) < \varepsilon/3$ and $g_n \to f$ uniformly on A, and there exist a compact $B \subset A$ and an open $U \supset E$ such that $\mu(A \setminus B) < \varepsilon/3$ and $\mu(U/E) < \varepsilon/3$. Since $g_n \to f$ uniformly on B, $f|B$ is continuous, so by the Tietze extension theorem (4.34), there exists $g \in C_c(X)$ such that $g = f$ on B and $\operatorname{supp}(g) \subset U$. But then $\{x: f(x) \neq g(x)\}$ is contained in $U \setminus B$, which has measure less than ε. Finally, let $\beta: \mathbf{C} \to \mathbf{C}$ be defined by $\beta(z) = z$ if $|z| \le \|f\|_u$ and $\beta(z) = \|f\|_u \operatorname{sgn} z$ if $|z| > \|f\|_u$. Then β is continuous, and $\varphi = \beta \circ g$ satisfies $\|\varphi\|_u \le \|f\|_u$ and $\varphi = f$ on the set where $g = f$, so we are done.

If f is unbounded, let $A_n = \{x: |f(x)| \le n\}$. Then A_n increases to E as $n \to \infty$, so $\mu(E \setminus A_n) < \varepsilon/2$ for sufficiently large n. By the above argument there exists $\varphi \in C_c(X)$ such that $\varphi = f\chi_{A_n}$ except on a set of measure $< \varepsilon/2$, and hence $\varphi = f$ except on a set of measure $< \varepsilon$. □

Our final group of results in this section concerns semi-continuous functions. If X is a topological space, a function $f: X \to (-\infty, \infty]$ is called **lower semi-continuous (l.s.c.)** if $\{x: f(x) > a\}$ is open for all $a \in \mathbf{R}$, and $f: X \to [-\infty, \infty)$ is called **upper semi-continuous (u.s.c.)** if $\{x: f(x) < a\}$ is open for all $a \in \mathbf{R}$.

(7.11) Proposition. Let X be a topological space.

(a) If U is open in X, then χ_U is l.s.c.

(b) If f is l.s.c. and $c \in [0, \infty)$, then cf is l.s.c.

(c) If $\mathscr{G}$ is a family of l.s.c. functions and $f(x) = \sup\{g(x): g \in \mathscr{G}\}$, then f is l.s.c.

(d) If f_1, f_2 are l.s.c. then so is $f_1 + f_2$.

(e) If X is an LCH space and f is l.s.c. and nonnegative, then

$$f(x) = \sup\{g(x): g \in C_c(X) \text{ and } 0 \le g \le f\}.$$

Proof. (a) and (b) are obvious, and (c) follows from the observation that $f^{-1}((a,\infty]) = \bigcup_{g\in\mathscr{G}} g^{-1}((a,\infty])$. As for (d), if $f_1(x_0)+f_2(x_0) > a$, choose $\varepsilon > 0$ so that $f_1(x_0) > a - f_2(x_0) + \varepsilon$. Then

$$\{x\colon f_1(x)+f_2(x) > a\} \supset \{x\colon f_1(x) > a - f_2(x_0) + \varepsilon\} \cap \{x\colon f_2(x) > f_2(x_0) - \varepsilon\},$$

which is a neighborhood of x_0. Thus $f_1 + f_2$ is l.s.c. Finally, if X is LCH, $f(x) > 0$, and $0 < a < f(x)$, then $U = \{y\colon f(y) > a\}$ is an open set containing x, so by Urysohn's lemma there exists $g \in C_c(X)$ such that $g(x) = a$ and $0 \le g \le a\chi_U \le f$. This establishes (e) when $f(x) \ne 0$, and (e) is trivial when $f(x) = 0$. □

There is, of course, a corresponding set of results for u.s.c. functions, whose formulation is left to the reader. The following result is a sort of monotone convergence theorem for nets.

(7.12) Proposition. Let $\mathscr{G}$ be a family of nonnegative l.s.c. functions on an LCH space X which is directed by $\le$ (that is, for every $g_1, g_2 \in \mathscr{G}$ there exists $g \in \mathscr{G}$ with $g_1 \le g$ and $g_2 \le g$). Let $f = \sup\{g\colon g \in \mathscr{G}\}$. If μ is any Radon measure on X, then $\int f\,d\mu = \sup\{\int g\,d\mu\colon g \in \mathscr{G}\}$.

Proof. By Proposition (7.11), f is l.s.c. and hence Borel measurable, and clearly $\int f\,d\mu \ge \sup\{\int g\,d\mu\}$. To prove the reverse inequality, consider the sequence $\{\varphi_n\}$ of simple functions increasing to f which was constructed in the proof of Theorem (2.10):

$$\varphi_n = 2^{-n}\sum_{j=1}^{2^{2n}} \chi_{U_{nj}}, \text{ where } U_{nj} = \{x\colon f(x) > j2^{-n}\}.$$

By the monotone convergence theorem, given $a < \int f\,d\mu$ we can fix n large enough so that $2^{-n}\sum_j \mu(U_{nj}) = \int \varphi_n\,d\mu > a$. Since U_{nj} is open, for $1 \le j \le 2^{2n}$ there is a compact $K_j \subset U_{nj}$ such that $2^{-n}\sum_j \mu(K_j) > a$. Let $\psi = 2^{-n}\sum_j \chi_{K_j}$. For each $x \in \bigcup_j K_j$ we have $f(x) > \varphi_n(x) \ge \psi(x)$, so we can pick $g_x \in \mathscr{G}$ such that $g_x(x) > \psi(x)$. But $-\chi_{K_j}$ is l.s.c., so $g_x - \psi$ is l.s.c. by Proposition (7.11d), and hence $V_x = \{y\colon \psi(y) < g_x(y)\}$ is open. Then $\{V_x\colon x \in \bigcup_j K_j\}$ is an open cover of $\bigcup_j K_j$, so there is a finite subcover $V_{x_1}, \dots, V_{x_m}$. Pick $g \in \mathscr{G}$ such that $g_{x_k} \le g$ for $k = 1, \dots, m$: then $\psi \le g$, so $\int g\,d\mu > a$. Since a was any number less than $\int f\,d\mu$, we are done. □

(7.13) Corollary. If f is nonnegative and l.s.c., then

$$\int f\,d\mu = \sup\left\{\int g\,d\mu\colon g \in C_c(X), 0 \le g \le f\right\}.$$

Proof. Combine Propositions (7.11e) and (7.12). □

(7.14) Proposition. If μ is a Radon measure on X and f is a nonnegative Borel measurable function, then

$$\int f\,d\mu = \inf\left\{\int g\,d\mu\colon g \ge f \text{ and } g \text{ is l.s.c.}\right\}.$$

If $\{x\colon f(x) > 0\}$ is σ-finite, then

$$\int f\,d\mu = \sup\left\{\int g\,d\mu\colon 0 \le g \le f \text{ and } g \text{ is u.s.c.}\right\}.$$

Proof. Let $\{\varphi_n\}$ be a sequence of nonnegative simple functions which increase pointwise to f. Then $f = \varphi_1 + \sum_2^\infty(\varphi_n - \varphi_{n-1})$, which we can rewrite as $f = \sum_1^\infty a_j\chi_{E_j}$ where $a_j > 0$. Given $\varepsilon > 0$, choose for each j an open $U_j \supset E_j$ with $\mu(U_j) \le \mu(E_j) + \varepsilon 2^{-j}$. Then $g = \sum_1^\infty a_j\chi_{U_j}$ is l.s.c. by Proposition (7.11), $g \ge f$, and $\int g\,d\mu \le \int f\,d\mu + \varepsilon$. This establishes the first assertion. For the second, if $a < \int f\,d\mu$, let N be large enough so that $\sum_1^N a_j\mu(E_j) > a$. Since the E_j's are σ-finite, by Proposition (7.5) there are compact sets $K_j \subset E_j$ such that $\sum_1^N a_j\mu(K_j) > a$. Thus if $g = \sum_1^N a_j\chi_{K_j}$, g is u.s.c., $g \le f$, and $\int g\,d\mu > a$. □

EXERCISES

7. Let μ be a Radon measure on X and $A \in \mathscr{B}_X$. If A is either open or σ-finite, the Borel measure μ_A defined by $\mu_A(E) = \mu(E \cap A)$ is a Radon measure.

8. If μ is a Radon measure and $f \in L^1(\mu)$, then $\nu(E) = \int_E f\,d\mu$ is a Radon measure. [Use Corollary (3.6).]

9. If μ is a Radon measure on X and $\varphi \in C(X, [0,\infty))$, then $\nu(E) = \int_E \varphi\,d\mu$ is a Radon measure. (Let ν' be the Radon measure associated to the functional $f \to \int f\varphi\,d\mu$. Then (i) $\nu(U) = \nu'(U)$ when U is open [apply Corollary (7.13) to $f\chi_U$], and (ii) ν is outer regular on all Borel sets. Conclusion: $\nu = \nu'$.)

10. Let ν be a Radon measure on the σ-compact space X and let f be a nonnegative Borel measurable function on X such that $\int_K f\,d\mu < \infty$ for all compact $K \subset X$. Then $\nu(E) = \int_E f\,d\mu$ is a Radon measure.

11. If μ is a Radon measure and $f \in L^1(\mu)$ is real-valued, for every $\varepsilon > 0$ there are an l.s.c. function g and a u.s.c. function h such that $h \le f \le g$ and $\int (g - h)\,d\mu < \varepsilon$.

12. If μ is a Radon measure on X such that $\mu(\{x\}) = 0$ for all $x \in X$, and $0 < \mu(A) < \infty$, then for any α such that $0 < \alpha < \mu(A)$ there is a Borel set $B \subset A$ such that $\mu(B) = \alpha$.

13. Let $X = \mathbf{R} \times \mathbf{R}_d$, where $\mathbf{R}_d$ denotes $\mathbf{R}$ with the discrete topology. If f is a function on X, let $f^y(x) = f(x, y)$.

(a) $f \in C_c(X)$ iff $f^y \in C_c(\mathbf{R})$ for all y and $f^y = 0$ for all but finitely many y.

(b) Define a positive linear functional on $C_c(X)$ by $I(f) = \sum_{y \in \mathbf{R}} \int f(x, y)\, dx$, and let μ be the associated Radon measure on X. Then $\mu(E) = \infty$ for all E such that $E \cap (\mathbf{R} \times \{y\}) \neq 0$ for uncountably many y.

(c) Let $E = \{0\} \times \mathbf{R}_d$. Then $\mu(E) = \infty$ but $\mu(K) = 0$ for all compact $K \subset E$.

14. Let μ be a Radon measure and let μ_0 be the semifinite part of μ (see Exercise 1.15).

(a) μ_0 is inner regular on all Borel sets.

(b) μ_0 is outer regular on all Borel sets E such that $\mu(E) < \infty$.

(c) $\int f d\mu = \int f d\mu_0$ for all $f \in C_c(X)$.

(d) If μ is the measure of Exercise 13, then $\mu_0(E) = \sum_{y \in \mathbf{R}} m(E \cap (\mathbf{R} \times \{y\}))$ for all Borel sets E, where m is Lebesgue measure on $\mathbf{R}$.

15. Let Ω be the set of countable ordinals, ω_1 the first uncountable ordinal, and $\Omega^* = \Omega \cup \{\omega_1\}$. Let Ω^* be endowed with the order topology (Exercise 4.9).

(a) Ω^* is a compact Hausdorff space. (Hint: Ω^* contains no infinite strictly decreasing sequences.)

(b) Ω is an open set in Ω^* which is not σ-compact.

(c) A subset E of Ω^* is uncountable iff for each $x \in \Omega$ there exists $y \in \Omega \cap E$ such that $x < y$.

(d) If $\{E_n\}$ is a sequence of uncountable closed sets in Ω^*, then $\bigcap_1^\infty E_n$ is uncountable. (If $\{x_j\}$ is an increasing sequence in Ω such that each E_n contains infinitely many x_j's, then $\lim_{j \to \infty} x_j$ exists and is in $\bigcap_1^\infty E_n$.)

(e) If $E \subset \mathscr{B}_{\Omega^*}$, then either $E \cup \{\omega_1\}$ or $E^c \cup \{\omega_1\}$ contains an uncountable closed set. (The set of all E satisfying the latter condition is a σ-algebra.)

(f) Define μ on $\mathscr{B}_{\Omega^*}$ by $\mu(E) = 1$ if $E \cup \{\omega_1\}$ contains an uncountable closed set, $\mu(E) = 0$ otherwise. Then μ is a measure, $\mu(\{\omega_1\}) = 0$, but $\mu(U) = 1$ for every open U containing ω_1. Thus μ is not regular.

(g) If $f \in C(\Omega^*)$, there exists $x \in \Omega$ such that $f(y) = f(\omega_1)$ for $y \geq x$. [If $E_n = \{x: |f(x) - f(\omega_1)| < n^{-1}\}$, then E_n^c is countable.]

(h) With μ as in (f), the Radon measure on Ω^* associated to the functional $f \to \int f d\mu$ is the point mass at ω_1.

3. THE DUAL OF $C_0(X)$

We recall that for any LCH space X, $C_0(X)$ is the uniform closure of $C_c(X)$ [Proposition (4.35)], and hence if μ is a Radon measure on X, the functional $I(f) = \int f d\mu$ extends continuously to $C_0(X)$ iff it is bounded with respect to the uniform norm. In view of the equation

$$\mu(X) = \sup\{\textstyle\int f d\mu\colon f \in C_c(X), 0 \leq f \leq 1\}$$

together with the fact that $|\int f d\mu| \leq \int |f| d\mu$, this happens precisely when $\mu(X) < \infty$, in which case $\mu(X)$ is the operator norm of I.

We have therefore identified the positive linear functionals on $C_0(X)$: they are given by integration against finite Radon measures. Our object in this section is to extend this result to give a complete description of $C_0(X)^*$. The key fact is that real linear functionals on $C_0(X, \mathbf{R})$ have a "Jordan decomposition."

(7.15) Lemma. If $I \in C_0(X, \mathbf{R})^*$, there exist positive functionals $I^+, I^- \in C_0(X, \mathbf{R})^*$ such that $I = I^+ - I^-$.

Proof. If $f \in C_0(X, [0, \infty))$, we define

$$I^+(f) = \sup\{I(g)\colon g \in C_0(X, \mathbf{R}), 0 \leq g \leq f\}.$$

Since $|I(g)| \leq \|I\| \|g\|_u \leq \|I\| \|f\|_u$ for $0 \leq g \leq f$, $I^+(f)$ is finite and bounded by $\|I\| \|f\|_u$. We claim that I^+ is the restriction to $C_0(X, [0, \infty))$ of a linear functional; the proof is much the same as the proof of the linearity of the integral in Section 2.3. Obviously $I^+(cf) = cI^+(f)$ if $c \in [0, \infty)$. Also, whenever $0 \leq g_1 \leq f_1$ and $0 \leq g_2 \leq f_2$ we have $0 \leq g_1 + g_2 \leq f_1 + f_2$, so that $I^+(f_1 + f_2) \geq I(g_1) + I(g_2)$, and it follows that $I^+(f_1 + f_2) \geq I^+(f_1) + I^+(f_2)$. On the other hand, if $0 \leq g \leq f_1 + f_2$, let $g_1 = \min(g, f_1)$ and $g_2 = g - g_1$. Then $0 \leq g_1 \leq f_1$ and $0 \leq g_2 \leq f_2$, so $I(g) = I(g_1) + I(g_2) \leq I^+(f_1) + I^+(f_2)$; therefore $I^+(f_1 + f_2) \leq I^+(f_1) + I^+(f_2)$. In short, $I^+(f_1 + f_2) = I^+(f_1) + I^+(f_2)$.

Now, if $f \in C_0(X, \mathbf{R})$, then $f^+, f^- \in C_0(X, [0, \infty))$, and we define $I^+(f) = I^+(f^+) - I^+(f^-)$. If also $f = g - h$ where $g, h \geq 0$, then $g + f^- = h + f^+$, whence $I^+(g) + I^+(f^-) = I^+(h) + I^+(f^+)$. Thus $I^+(f) = I^+(g) - I^+(h)$, and it follows easily that I^+ is linear on $C_0(X, \mathbf{R})$. Moreover,

$$|I^+(f)| \leq \max(I^+(f^+), I^+(f^-)) \leq \|I\| \max(\|f^+\|_u, \|f^-\|_u) = \|I\| \|f\|_u,$$

so that $\|I^+\| \leq \|I\|$.

Finally, let $I^- = I^+ - I$. Then $I^- \in C_0(X, \mathbf{R})^*$, and it is immediate from the definition of I^+ that I^+ and I^- are positive. □

Any $I \in C_0(X)^*$ is uniquely determined by its restriction J to $C_0(X, \mathbf{R})$, and we have $J = J_1 + iJ_2$ where J_1, J_2 are real linear functionals. We therefore conclude from Lemma (7.15) and the discussion preceding it that for any $I \in C_0(X)^*$ there are finite Radon measures μ_j $(1 \le j \le 4)$ such that $I(f) = \int f d\mu$ where $\mu = \mu_1 - \mu_2 + i(\mu_3 - \mu_4)$.

At this point we need some more definitions. A **signed Radon measure** is a signed Borel measure whose positive and negative variations are Radon measures, and a **complex Radon measure** is a complex Borel measure whose real and imaginary parts are signed Radon measures. [It is worth noting that on a second countable LCH space, every complex Borel measure is Radon. This follows from Theorem (7.8) since complex measures are bounded.] We denote the space of complex Radon measures on X by $M(X)$, and for $\mu \in M(X)$ we define

$$\|\mu\| = |\mu|(X),$$

where, of course, $|\mu|$ is the total variation of μ.

(7.16) Proposition. If μ is a complex Borel measure, then μ is Radon iff $|\mu|$ is Radon. Moreover, $M(X)$ is a vector space and $\mu \to \|\mu\|$ is a norm on it.

Proof. We observe that a finite positive Borel measure ν is Radon iff for every Borel set E and every $\varepsilon > 0$ there exist a compact K and an open U such that $K \subset E \subset U$ and $\nu(U \setminus K) < \varepsilon$ [by Propositions (7.5) and (7.7)]. (a) follows easily from this: if $\mu = \mu_1 - \mu_2 + i(\mu_3 - \mu_4)$ and $|\mu|(U \setminus K) < \varepsilon$, then $\mu_j(U \setminus K) < \varepsilon$ for all j; conversely, if $\mu_j(U_j \setminus K_j) < \varepsilon/4$ for $1 \le j \le 4$ and $K = \bigcup_1^4 K_j$, $U = \bigcap_1^4 U_j$, then $|\mu|(U \setminus K) < \varepsilon$. The same sort of argument shows that $M(X)$ is closed under addition and scalar multiplication. That $\| \ \|$ is a norm on $M(X)$ follows from Proposition (3.14). □

(7.17) The Riesz Representation Theorem. Let X be an LCH space, and for $\mu \in M(X)$ and $f \in C_0(X)$ let $I_\mu(f) = \int f d\mu$. Then the map $\mu \to I_\mu$ is an isometric isomorphism from $M(X)$ to $C_0(X)^*$.

Proof. We have already shown that every $I \in C_0(X)^*$ is of the form I_μ. On the other hand, if $\mu \in M(X)$, by Proposition (3.17c),

$$\left|\int f d\mu\right| \le \int |f| d|\mu| \le \|f\|_u \|\mu\|,$$

so $I_\mu \in C_0(X)^*$ and $\|I_\mu\| \le \|\mu\|$. Moreover, if $h = d\mu/d|\mu|$, then $|h| = 1$, so by Lusin's theorem, for any $\varepsilon > 0$ there exists $f \in C_c(X)$ such that $\|f\|_u \le 1$ and $f = \bar{h}$ except on a set E with $|\mu|(E) < \varepsilon/2$. Then

$$\|\mu\| = \int |h|^2 d|\mu| = \int \bar{h} d\mu \le \left|\int f d\mu\right| + \left|\int (f - \bar{h}) d\mu\right|$$

$$\le \left|\int f d\mu\right| + 2|\mu|(E) < \left|\int f d\mu\right| + \varepsilon.$$

It follows that $\|\mu\| \le \|I_\mu\|$, so the proof is complete. □

(7.18) Corollary. If X is a compact Hausdorff space, then $C(X)^*$ is isometrically isomorphic to $M(X)$.

Let μ be a fixed positive Radon measure on X. If $f \in L^1(\mu)$, the complex measure $d\nu_f = f d\mu$ is easily seen to be Radon (Exercise 8), and $\|\nu_f\| = \int |f| d\mu = \|f\|_1$. Thus $f \to \nu_f$ is an isometric embedding of $L^1(\mu)$ into $M(X)$ whose range consists precisely of those $\nu \in M(X)$ such that $\nu \ll \mu$. (The last statement follows from the Radon–Nikodym theorem, which applies even if μ is not σ-finite: see Section 5.) The most important example of this situation is $\mu = m$ = Lebesgue measure on $\mathbf{R}^n$, and we shall identify $L^1(m)$ with a subspace of $M(\mathbf{R}^n)$.

The weak * topology on $M(X) \cong C_0(X)^*$ is of considerable importance in applications; we shall call it the **vague topology** on $M(X)$. [The term "vague" is common in probability theory and has the advantage of forming an adverb more gracefully than "weak *." The vague topology is sometimes called the weak topology, but this terminology conflicts with ours, since $C_0(X)$ is rarely reflexive.] In the context of the preceding paragraph, weak convergence arguments for $L^p(\mu)$ generally fail for $p = 1$ because $L^1(\mu)$ is not the dual of $L^\infty(\mu)$, but good substitute results can often be obtained by regarding $L^1(\mu)$ as a subspace of $M(X)$ and using the vague topology there. We conclude by presenting a useful criterion for vague convergence in $M(\mathbf{R})$.

(7.19) Proposition. Suppose $\mu_n, \mu \in M(\mathbf{R})$ $(n \in \mathbf{N})$, and let $F_n(x) = \mu_n((-\infty, x])$, $F(x) = \mu((-\infty, x])$.

(a) If $\sup_n \|\mu_n\| < \infty$ and $F_n(x) \to F(x)$ for every x at which F is continuous, then $\mu_n \to \mu$ vaguely.

(b) The converse of (a) is true provided the μ_n's are positive.

Proof. (a) Since F is continuous except at countably many points [Theorems (3.27) and (3.29)], $F_n \to F$ a.e. with respect to Lebesgue measure. Also, $\|F_n\|_u \le \|\mu_n\|$, so the F_n's are uniformly bounded. If f is continuously differentiable and has compact support, then, integration by parts [Theorem (3.30)] and the dominated convergence theorem yield

$$\int f d\mu_n = \int f'(x) F_n(x) dx \to \int f'(x) F(x) dx = \int f d\mu.$$

But by the Stone–Weierstrass theorem (4.52), the set of all such f's is dense in $C_0(\mathbf{R})$, so $\int f d\mu_n \to \int f d\mu$ for all $f \in C_0(\mathbf{R})$ by Proposition (5.17). This means that $\mu_n \to \mu$ vaguely.

(b) If $\mu_n \to \mu$ vaguely, then $\sup_n \|\mu_n\| < \infty$ by the uniform boundedness principle. Suppose that $\mu_n \ge 0$, and hence $\mu \ge 0$, and that F is continuous at $x = a$. If $f \in C_0(\mathbf{R})$ is the function which is 1 on $[-N, a]$, 0 on $(-\infty, -N - \varepsilon]$

and $[a + \varepsilon, \infty)$, and linear in between, we have

$$F_n(a) - F_n(-N) = \mu_n((-N, a]) \leq \int f d\mu_n \to \int f d\mu$$

$$\leq F(a + \varepsilon) - F(-N - \varepsilon).$$

As $N \to \infty$, $F_n(-N)$ and $F(-N - \varepsilon)$ tend to zero, so

$$\limsup_{n \to \infty} F_n(a) \leq F(a + \varepsilon).$$

Similarly by considering the function which is 1 on $[-N + \varepsilon, a - \varepsilon]$, 0 on $(-\infty, N]$ and $[a, \infty)$, and linear in between, we see that

$$\liminf_{n \to \infty} F_n(a) \geq F(a - \varepsilon).$$

Since ε is arbitrary and F is continuous at a, the proof is complete. □

EXERCISES

16. Suppose $I \in C_0(X, \mathbf{R})^*$ and I^+, I^- are the functionals constructed in the proof of Lemma (7.15). If μ is the signed Radon measure associated to I, then the positive and negative variations of μ are the Radon measures associated to I^+ and I^-.

17. If μ is a positive Radon measure on X with $\mu(X) = \infty$, there exists $f \in C_0(X)$ such that $\int f d\mu = \infty$. Consequently, every positive linear functional on $C_0(X)$ is bounded.

18. If μ is a σ-finite Radon measure on X and $\nu \in M(X)$, let $\nu = \nu_1 + \nu_2$ be the Lebesgue decomposition of ν with respect to μ. Then ν_1 and ν_2 are Radon. (Use Exercise 8.)

19. Let X be a completely regular space and $\mathscr{A}$ a completely regular subalgebra of $BC(X)$ (see Exercise 4.73). Find a description of $\mathscr{A}^*$ as a space of measures.

20. Some examples of non-reflexivity of $C_0(X)$:

(a) If $\mu \in M(X)$, let $\Phi(\mu) = \sum_{x \in X} \mu(\{x\})$. This sum is well defined, and $\Phi \in M(X)^*$. If there exists a nonzero $\mu \in M(X)$ such that $\mu(\{x\}) = 0$ for all $x \in X$, then Φ is not in the image of $C_0(X)$ in $M(X)^* \cong C_0(X)^{**}$.

(b) At the other extreme, let $X = \mathbf{N}$ with the discrete topology; $C_0(X)$ is then usually denoted by c_0. We have $(c_0)^* \cong l^1$ and $(l^1)^* \cong l^\infty$.

21. Let $\{f_\alpha\}_{\alpha \in A}$ be a subset of $C_0(X)$ and $\{c_\alpha\}_{\alpha \in A}$ a family of complex numbers. If for each finite subset B of A there exists $\mu_B \in M(X)$ such that $\|\mu_B\| \leq 1$ and $\int f_\alpha d\mu_B = c_\alpha$ for $\alpha \in B$, then there exists $\mu \in M(X)$ such that $\|\mu\| \leq 1$ and $\int f_\alpha d\mu = c_\alpha$ for all $\alpha \in A$.

22. A sequence $\{f_n\}$ in $C_0(X)$ converges weakly to $f \in C_0(X)$ iff $\sup_n \|f_n\|_u < \infty$ and $f_n \to f$ pointwise.

23. The hypothesis of positivity in Proposition (7.19b) is necessary. (Take μ_n to be the difference of the point masses at n^{-1} and $-n^{-1}$.)

24. Find examples of sequences $\{\mu_n\}$ in $M(\mathbf{R})$ such that:

(a) $\mu_n \to 0$ vaguely, but $\|\mu_n\| \not\to 0$.

(b) $\mu_n \to 0$ vaguely, but $\int f d\mu_n \not\to \int f d\mu$ for some bounded measurable f with compact support.

(c) $\mu_n \geq 0$ and $\mu_n \to \mu$ vaguely, but, in the notation of Proposition (7.19), there exists $x \in \mathbf{R}$ such that $F_n(x) \not\to F(x)$.

25. Let μ be a Radon measure on X such that every open set has positive measure (e.g., Lebesgue measure). For each $x \in X$ there is a sequence $\{f_n\}$ in $L^1(\mu)$ which converges vaguely in $M(X)$ to the point mass at x. [Take $f_n = \mu(U_n)^{-1} \chi_{U_n}$ for a suitable sequence $\{U_n\}$.]

26. If $\{\mu_n\} \subset M(X)$, $\mu_n \to \mu$ vaguely, and $\|\mu_n\| \to \|\mu\|$, then $\int f d\mu_n \to \int f d\mu$ for every $f \in BC(X)$. [If $\mu = 0$ the result is trivial. Otherwise, pick $g \in C_c(X)$ such that $\int g d\mu > \|\mu\| - \varepsilon$: then $\int gf d\mu_n \to \int gf d\mu$ for $f \in BC(X)$.] Moreover, the hypothesis $\|\mu_n\| \to \|\mu\|$ cannot be omitted.

27. Let $C^k([0, 1])$ be as in Exercise 5.7. If $I \in C^k([0, 1])^*$, there exist a unique complex Radon measure μ on $[0, 1]$ and constants $c_0, \ldots, c_{k-1}$ such that

$$I(f) = \int f^{(k)} d\mu + \sum_0^{k-1} c_j f^{(j)}(0).$$

[One could replace the functionals $f \to f^{(j)}(0)$ by any set of k functionals which separate points in the space of polynomials of degree $\leq k - 1$.]

4. PRODUCTS OF RADON MEASURES

In this section we study Radon measures on product spaces. X and Y will denote LCH spaces, and π_X and π_Y will denote the projections of $X \times Y$ onto X and Y, respectively.

(7.20) Theorem.

(a) $\mathscr{B}_X \otimes \mathscr{B}_Y \subset \mathscr{B}_{X \times Y}$.

(b) If X and Y are second countable, $\mathscr{B}_X \otimes \mathscr{B}_Y = \mathscr{B}_{X \times Y}$.

(c) If X and Y are second countable and μ, ν are Radon measures on X, Y respectively, then $\mu \times \nu$ is a Radon measure on $X \times Y$.

Proof. Parts (a) and (b) are direct generalizations of Proposition (1.5), and the proof is the same. The essential tool is Proposition (1.4). It implies, first, that $\mathscr{B}_X \otimes \mathscr{B}_Y$ is generated by the sets $U \times V$ where U is open in X and V is open in Y. Since these sets are open in $X \times Y$, we have $\mathscr{B}_X \otimes \mathscr{B}_Y \subset \mathscr{B}_{X\times Y}$. If the topologies on X and Y have countable bases $\mathscr{E}$ and $\mathscr{F}$, then every open set in X, Y, and $X \times Y$ is a countable union of sets in $\mathscr{E}$, $\mathscr{F}$, and $\{U \times V: U \in \mathscr{E}, V \in \mathscr{F}\}$. It follows that $\mathscr{B}_X$, $\mathscr{B}_Y$, and $\mathscr{B}_{X\times Y}$ are generated by these families and hence that $\mathscr{B}_X \otimes \mathscr{B}_Y = \mathscr{B}_{X\times Y}$. As for (c), $\mu \times \nu$ is a Borel measure by (b), and by Theorem (7.8) we need only show that $\mu \times \nu(K)$ is finite for any compact $K \subset X \times Y$. But this is easy: $\pi_X(K)$ and $\pi_Y(K)$ are compact, and $K \subset \pi_X(K) \times \pi_Y(K)$, so

$$\mu \times \nu(K) \leq \mu(\pi_X(K))\nu(\pi_Y(K)) < \infty. \qquad \square$$

When X or Y is not second countable it can happen that $\mathscr{B}_X \otimes \mathscr{B}_Y \neq \mathscr{B}_{X\times Y}$: see Exercises 28 and 29. In this case the product of Radon measures is certainly not a Radon measure. However, there is a natural way of manufacturing a Radon measure from it. To see this, we need a couple of facts about continuous functions. If f and g are functions on X and Y, we define $f \otimes g$ on $X \times Y$ by

$$f \otimes g(x, y) = f(x)g(y).$$

(7.21) Proposition. Let $\mathscr{P}$ be the linear span of $\{f \otimes g: f \in C_c(X), g \in C_c(Y)\}$. Then $\mathscr{P}$ is dense in $C_c(X \times Y)$ in the uniform norm. More precisely, if $h \in C_c(X \times Y)$, $\varepsilon > 0$, and U and V are precompact open sets in X and Y containing $\pi_X(\text{supp}(h))$ and $\pi_Y(\text{supp}(h))$, there exists $F \in \mathscr{P}$ such that $\|F - h\|_u < \varepsilon$ and $\text{supp}(F) \subset U \times V$.

Proof. $\overline{U} \times \overline{V}$ is a compact Hausdorff space. It follows easily from the Stone–Weierstrass theorem that the linear span of $\{f \otimes g: f \in C(\overline{U}), g \in C(\overline{V})\}$ is dense in $C(\overline{U} \times \overline{V})$. In particular there is an element H of this linear span such that $\sup_{\overline{U}\times\overline{V}}|H - h| < \varepsilon$. Next, by Urysohn's lemma there exist $\varphi \in C_c(U, [0,1])$ and $\psi \in C_c(V, [0,1])$ such that $\varphi = 1$ on $\pi_X(\text{supp}(h))$ and $\psi = 1$ on $\pi_Y(\text{supp}(h))$. Then if we define $F = (\varphi \otimes \psi)H$ on $\overline{U} \times \overline{V}$ and $F = 0$ elsewhere, we have $F \in \mathscr{P}$, $\text{supp}(F) \subset \overline{U} \times \overline{V}$, and $\|F - h\|_u < \varepsilon$. $\square$

(7.22) Proposition. Every $f \in C_c(X \times Y)$ is $\mathscr{B}_X \otimes \mathscr{B}_Y$-measurable. Moreover, if μ, ν are Radon measures on X, Y, then $C_c(X \times Y) \subset L^1(\mu \times \nu)$, and if $f \in C_c(X \times Y)$, then $\int f d(\mu \times \nu) = \int\int f d\mu\, d\nu = \int\int f d\nu d\mu$.

Proof. If $f \in C_c(X)$ and $g \in C_c(Y)$, we have $f \otimes g = (f \circ \pi_X)(g \circ \pi_Y)$. Since π_X and π_Y are measurable from $\mathscr{B}_X \otimes \mathscr{B}_Y$ to $\mathscr{B}_X$ and $\mathscr{B}_Y$ (by definition of $\mathscr{B}_X \otimes \mathscr{B}_Y$), and f and g are continuous, $f \circ \pi_X$ and $g \circ \pi_Y$ are $\mathscr{B}_X \otimes \mathscr{B}_Y$-measurable. Since products, sums, and pointwise limits of measurable functions are measurable, the first assertion follows from Proposition (7.21). Next, every $f \in C_c(X \times Y)$ is bounded and supported in a set of finite $\mu \times \nu$-measure, hence is in $L^1(\mu \times \nu)$. Fubini's theorem holds for such f even if μ and ν are not σ-finite because one can replace μ and ν by the finite measures $\mu|\pi_X(\text{supp}(f))$ and $\nu|\pi_Y(\text{supp}(f))$. $\square$

It is now clear how to obtain a Radon measure on $X \times Y$ from Radon measures μ and ν on X and Y. Namely, by Proposition (7.22) the formula $I(f) = \int f d(\mu \times \nu)$ defines a positive linear functional on $C_c(X \times Y)$, so it determines a Radon measure on $X \times Y$ by the Riesz representation theorem. We call this measure the **Radon product** of μ and ν and denote it by $\mu \hat{\times} \nu$.

The obvious question is: does $\mu \hat{\times} \nu$ agree with $\mu \times \nu$ on $\mathscr{B}_X \otimes \mathscr{B}_Y$? In general, the answer is no. Indeed, a counterexample may be obtained by taking $X = \mathbf{R}$, $Y = \mathbf{R}_d$ ($\mathbf{R}$ with the discrete topology), $\mu =$ Lebesgue measure, and $\nu =$ counting measure. It is not hard to see that in this case $\mathscr{B}_{X\times Y} = \mathscr{B}_X \otimes \mathscr{B}_Y$, but Exercises 13 and 14 show that $\mu \hat{\times} \nu$ is not semifinite and that $\mu \times \nu$ is the semifinite part of $\mu \hat{\times} \nu$. However, some results are still available, and in the σ-finite case everything works out beautifully. In what follows, we employ the notation of x-sections and y-sections introduced in Section 2.5.

(7.23) Lemma.

(a) If $E \in \mathscr{B}_{X\times Y}$, then $E_x \in \mathscr{B}_Y$ for all $x \in X$ and $E^y \in \mathscr{B}_X$ for all $y \in Y$.

(b) If $f: X \times Y \to \mathbf{C}$ is $\mathscr{B}_{X\times Y}$-measurable, then f_x is $\mathscr{B}_Y$-measurable for all $x \in X$ and f^y is $\mathscr{B}_X$-measurable for all $y \in Y$.

Proof. The collection of all $E \subset X \times Y$ such that $E_x \in \mathscr{B}_Y$ and $E^y \in \mathscr{B}_X$ for all x, y is easily seen to be a σ-algebra. It contains all open sets [if E is open, so are E_x and E^y, being inverse images of E under the maps $y' \to (x, y')$ and $x' \to (x', y)$], hence it contains $\mathscr{B}_{X\times Y}$. This proves (a), and (b) follows since $(f_x)^{-1}(A) = (f^{-1}(A))_x$ and $(f^y)^{-1}(A) = (f^{-1}(A))^y$. $\square$

(7.24) Lemma. If $f \in C_c(X \times Y)$ and μ, ν are Radon measures on X, Y, then $\int f_x\, d\nu$ and $\int f^y\, d\mu$ are continuous functions of x and y.

Proof. We write out the proof only for f_x. It suffices to show that for any $x_0 \in X$ and $\varepsilon > 0$ there exists a neighborhood U of x_0 such that $\|f_x - f_{x_0}\|_u < \varepsilon$ for $x \in U$, since then

$$\left|\int (f_x - f_{x_0})\, d\nu\right| \leq \varepsilon\nu(\pi_Y(\text{supp}(f))).$$

However, for each $y \in \pi_Y(\text{supp}(f))$ there exist neighborhoods U_y, V_y of x_0 and y such that if $(x, z) \in U_y \times V_y$, then $|f(x_0, y) - f(x, z)| < \varepsilon/2$. We may choose a finite subcover $V_{y_1}, \ldots, V_{y_m}$ of $\pi_Y(\text{supp}(f))$ and then take $U = \bigcap_1^m U_{y_j}$; details are left to the reader. $\square$

(7.25) Proposition. Let μ and ν be Radon measures on X and Y. If U is open in $X \times Y$, then $x \to \nu(U_x)$ and $y \to \mu(U_y)$ are Borel measurable on X and Y, and

$$\mu \hat{\times} \nu(U) = \int \nu(U_x)\, d\mu(x) = \int \mu(U^y)\, d\nu(y).$$

Proof. Let $\mathscr{F} = \{f \in C_c(X \times Y): 0 \le f \le \chi_U\}$. By Proposition (7.11) we have $\chi_U = \sup\{f: f \in \mathscr{F}\}$ and hence $\chi_{U_x} = \sup\{f_x: f \in \mathscr{F}\}$ and $\chi_{U^y} = \sup\{f^y: f \in \mathscr{F}\}$. Thus by Proposition (7.12),

$$\mu \hat{\times} \nu(U) = \sup\left\{\int f d(\mu \hat{\times} \nu): f \in \mathscr{F}\right\},$$

$$\nu(U_x) = \sup\left\{\int f_x d\nu: f \in \mathscr{F}\right\},\ \mu(U^y) = \sup\left\{\int f^y d\mu: f \in \mathscr{F}\right\}.$$

From Lemma (7.24) and Proposition (7.11) it follows that $x \to \nu(U_x)$ and $y \to \mu(U^y)$ are l.s.c. and hence Borel measurable. Another application of Proposition (7.12), together with Proposition (7.22), yields

$$\mu \hat{\times} \nu(U) = \sup\left\{\int\int f_x\, d\nu\, d\mu(x): f \in \mathscr{F}\right\}$$

$$= \int \left(\sup\left\{\int f_x d\nu: f \in \mathscr{F}\right\}\right) d\mu(x) = \int \nu(U_x)\, d\mu(x),$$

and likewise $\mu \hat{\times} \nu(U) = \int \mu(U^y) d\nu(y)$. □

(7.26) Theorem. Suppose μ and ν are σ-finite Radon measures on X and Y. If $E \in \mathscr{B}_{X\times Y}$, then the functions $x \to \nu(E_x)$ and $y \to \mu(E^y)$ [which make sense by Lemma (7.23)] are Borel measurable on X and Y, and

$$\mu \hat{\times} \nu(E) = \int \nu(E_x)\, d\mu(x) = \int \mu(E^y)\, d\nu(y).$$

Moreover, the restriction of $\mu \hat{\times} \nu$ to $\mathscr{B}_X \otimes \mathscr{B}_Y$ is $\mu \times \nu$.

Proof. For the moment, let us fix open sets $U \subset X$ and $V \subset Y$ with $\mu(U)$ and $\nu(V)$ finite, and let $W = U \times V$. Let $\mathscr{M}$ be the collection of all sets $E \in \mathscr{B}_{X\times Y}$ such that $E \cap W$ satisfies the conclusions of the theorem. We then have:

(i) $\mathscr{M}$ contains all open sets [by Proposition (7.25)].

(ii) If $E, F \in \mathscr{M}$ and $F \subset E$, then $E \setminus F \in \mathscr{M}$; in particular, if $F \in \mathscr{M}$, then $F^c = X \setminus F \in \mathscr{M}$. Indeed, we have

$$\mu \hat{\times} \nu(E \cap W) = \mu \hat{\times} \nu(F \cap W) + \mu \hat{\times} \nu((E \setminus F) \cap W),$$

and likewise for $\nu((E \cap W)_x)$ and $\mu((E \cap W)^y)$. Since the conclusions are true for $E \cap W$ and $F \cap W$ and all quantities involved are finite (this is why we introduced W), we can subtract to obtain the conclusions for $(E \setminus F) \cap W$.

(iii) $\mathscr{M}$ is closed under finite disjoint unions. (This is simply the additivity of the measures.)

(iv) $\mathscr{M}$ is closed under countable increasing unions, and hence [by (ii)] under countable decreasing intersections. (This follows from the monotone convergence theorem.)

Now let $\mathscr{E} = \{A \setminus B: A, B \text{ open in } X \times Y\}$ and let $\mathscr{A}$ be the collection of finite disjoint unions of sets in $\mathscr{E}$. Since

$$(A_1 \setminus B_1) \cap (A_2 \setminus B_2) = (A_1 \cup A_2) \setminus (B_1 \cup B_2),$$

$$(A \setminus B)^c = [(X \times Y) \setminus A] \cup [(A \cap B) \setminus \varnothing],$$

$\mathscr{E}$ is an elementary family, so by Proposition (1.7), $\mathscr{A}$ is an algebra. Therefore, by Lemma (2.35) the monotone class generated by $\mathscr{A}$ coincides with the σ-algebra generated by $\mathscr{A}$, which is clearly $\mathscr{B}_{X\times Y}$. But by (i)–(iv) [since $A \setminus B = A \setminus (A \cap B)$], $\mathscr{M}$ contains this monotone class, so $\mathscr{M} = \mathscr{B}_{X\times Y}$.

Next, since μ and ν are σ-finite and outer regular, we have $X = \bigcup_1^\infty U_n$, $Y = \bigcup_1^\infty V_n$ where U_n, V_n are open and have finite measure, and we may assume that $U_n \subset U_{n+1}$, $V_n \subset V_{n+1}$ for all n. The preceding argument (with $W = U_n \times V_n$) and the monotone convergence theorem then show that every $E \in \mathscr{B}_{X\times Y}$ satisfies the conclusions of the theorem.

Finally, if $E \in \mathscr{B}_X \otimes \mathscr{B}_Y$, by Tonelli's theorem we have

$$\mu \times \nu(E) = \int \nu(E_x)\, d\mu(x) = \mu \hat{\times} \nu(E),$$

and the proof is complete. □

(7.27) The Fubini–Tonelli Theorem for Radon Products. Let μ and ν be σ-finite Radon measures on X and Y, and let f be a Borel measurable function on $X \times Y$.

(a) If $f \ge 0$, then $x \to \int f_x d\nu$ and $y \to \int f^y d\mu$ [which make sense by Lemma (7.23)] are Borel measurable on X and Y.

(b) If $f \in L^1(\mu \hat{\times} \nu)$, then $f_x \in L^1(\nu)$ for a.e. x, $f^y \in L^1(\mu)$ for a.e. y, and $x \to \int f_x d\nu$ and $y \to \int f^y d\mu$ are in $L^1(\mu)$ and $L^1(\nu)$.

In both cases (a) and (b), we have

$$\int f d(\mu \hat{\times} \nu) = \int\int f\, d\mu\, d\nu = \int\int f\, d\nu\, d\mu.$$

The proof is identical to that of the ordinary Fubini–Tonelli theorem, except that Theorem (7.26) is used in place of Theorem (2.36).

EXERCISES

28. If X is the set of ordinals less than or equal to the first uncountable ordinal ω_1, with the order topology, then $\mathscr{B}_{X\times X} \neq \mathscr{B}_X \otimes \mathscr{B}_X$. In fact, $\{(x, y)\colon x < y < \omega_1\}$ is open but is not in $\mathscr{B}_X \otimes \mathscr{B}_X$. (Reexamine Exercise 2.47 in the light of Exercise 7.15.)

29. If X is a set of cardinality $> \mathfrak{c}$ with the discrete topology, then $\mathscr{B}_{X\times X} \neq \mathscr{B}_X \otimes \mathscr{B}_X$. In fact, $D = \{(x, y)\colon x = y\}$ is closed but not in $\mathscr{B}_X \otimes \mathscr{B}_X$. [Use Exercise 1.5 and Proposition (1.23). If $D \in \mathscr{B}_X \otimes \mathscr{B}_X$, then $D \in \mathscr{M}$ where $\mathscr{M}$ is a sub-σ-algebra of $\mathscr{B}_X \otimes \mathscr{B}_X$ generated by a countable family of rectangles, hence $D \in \mathscr{N} \otimes \mathscr{N}$ where $\mathscr{N}$ is a countably generated sub-σ-algebra of $\mathscr{B}_X$. Then $\{x\} = D_x \in \mathscr{N}$ for all x, but $\operatorname{card}(\mathscr{N}) \leq \mathfrak{c}$.] The same reasoning applies if X is replaced by its one-point compactification.

30. Let μ, ν be Radon measures on X, Y, not necessarily σ-finite. If f is a nonnegative l.s.c. function on $X \times Y$, then $x \to \int f_x d\nu$ and $y \to \int f^y d\mu$ are Borel measurable and $\int f d(\mu \hat{\times} \nu) = \int\int f\, d\mu\, d\nu = \int\int f\, d\nu\, d\mu$.

31. Some results concerning Baire sets on product spaces:

(a) $\mathscr{B}^0_{X\times Y} \subset \mathscr{B}^0_X \otimes \mathscr{B}^0_Y$. [Hint: Proposition (7.22) remains true if $\mathscr{B}$ is replaced by $\mathscr{B}^0$.]

(b) If X and Y are either compact or second countable, then $\mathscr{B}^0_{X\times Y} = \mathscr{B}^0_X \otimes \mathscr{B}^0_Y$.

(c) If X is an uncountable set with the discrete topology, then $\mathscr{B}^0_{X\times X} \neq \mathscr{B}^0_X \otimes \mathscr{B}^0_X$.

5. NOTES AND REFERENCES

The Riesz representation theorem is actually the work of many hands. F. Riesz [82] first proved it for the case $X = [a, b]$ $(a, b \in \mathbf{R})$; he formulated the result in terms of Riemann–Stieltjes integrals and used no measure theory. It was extended to compact subsets of $\mathbf{R}^n$ by Radon [77], to compact metric spaces by Banach (see Saks [93]) and to compact Hausdorff spaces by Kakutani [55]. A version of the theorem in the locally compact case was apparently known to Bourbaki about 1940 (see Weil [109], Section 6), but his treatment of integration theory was not published until 1952, by which time several others had obtained similar results. For more detailed references, see Dunford and Schwartz [26], Section IV.16, and Hewitt and Ross [53], Section 11.

Our use of the term "Radon measure," which derives from Radon's seminal paper [77], is common but not entirely standard. Some authors refer to such measures as "regular Borel measures," and others use the term "Radon measure" to mean a positive linear functional on $C_c(X)$. It should also be noted that some older texts define the Borel σ-algebra to be the σ-algebra generated by the compact sets, which is in general smaller than our $\mathscr{B}_X$.

If μ is a Radon measure, let $\bar{\mu}$ denote the complete, saturated extension of μ discussed at the end of Section 1. It is a significant fact that $\bar{\mu}$ is always decomposable in the sense of Exercise 3.15: see Hewitt and Stromberg [54], Theorem 19.30. Consequently, the extension of the Radon–Nikodym theorem in that exercise and the fact that $L^1(\bar{\mu})^* \cong \mathscr{L}^\infty(\bar{\mu})$ (Exercise 6.26) are available. In this connection one should also note that $L^1(\bar{\mu})$ is essentially identical to $L^1(\mu)$, by Propositions (2.12) and (2.20).

Section 2. Theorem (7.8) can be proved without involving the Riesz representation theorem: see Cohn [19], Proposition 7.2.3. Propositions (7.12) and (7.14) suggest an alternative way of constructing the Radon measure μ associated to a positive linear functional I on $C_c(X)$ in the spirit of Daniell integral (see Section 2.7). Namely, one first extends I to nonnegative l.s.c. functions g by setting

$$I(g) = \sup\{I(f)\colon f \in C_c(X), 0 \leq f \leq g\}$$

and then extends I to arbitrary nonnegative functions h by setting

$$I(h) = \inf\{I(g)\colon g \text{ is l.s.c. and } g \geq h\}.$$

It is then not difficult to verify that if $E \subset X$, $I(\chi_E) = \mu^*(E)$ where μ^* is the outer measure in the proof of Theorem (7.2). For details, see Hewitt and Ross [53] or Hewitt and Stromberg [54].

Section 3. Theorem (7.17) is frequently stated only for the case when X is compact; however, the more general formulation follows easily from the compact case by considering the one-point compactification of X. An interesting proof of the Baire measure version of this result, quite different from ours, can be found in Hartig [46].

Section 4. The Fubini–Tonelli theorem for Radon products, as presented here, is essentially due to deLeeuw [22]; see also Cohn [19], Section 7.6. Another variant of this theorem, which includes some further results for the non-σ-finite case, can be found in Hewitt and Ross [53], Section 13.

8

TOPICS IN FOURIER ANALYSIS

It is easy to say that Fourier analysis, or harmonic analysis, originated in the work of Euler, Fourier, and others on trigonometric series; it is much harder to describe succinctly what the subject comprises today, for it is a meeting ground for ideas from many parts of analysis and has applications in such diverse areas as partial differential equations and algebraic number theory. Two of the central themes of harmonic analysis, however, are the Fourier transform and convolution operators, which we study in this chapter along with the rudiments of distribution theory. In Section 7 we give a small sample of the many applications of these ideas.

1. PRELIMINARIES

We begin by making some notational conventions. Throughout this chapter we shall be working on $\mathbf{R}^n$, and n will always refer to the dimension. In any measure-theoretic considerations we always have Lebesgue measure in mind unless we specify otherwise. Thus, if E is a measurable set in $\mathbf{R}^n$, we shall denote $L^p(E, m)$ by $L^p(E)$, If U is open in $\mathbf{R}^n$ and $k \in \mathbf{N}$, we denote by $C^k(U)$ the space of all functions on U possessing continuous partial derivatives of order $\leq k$, and we set $C^\infty(U) = \bigcap_1^\infty C^k(U)$. Furthermore, for any $E \subset \mathbf{R}^n$ we denote by $C_c^\infty(E)$ the space of all C^∞ functions on $\mathbf{R}^n$ whose support is compact and contained in E. If $E = \mathbf{R}^n$ or $U = \mathbf{R}^n$, we shall usually omit it in naming function spaces: thus, $L^p = L^p(\mathbf{R}^n)$, $C^k = C^k(\mathbf{R}^n)$, $C_c^\infty = C_c^\infty(\mathbf{R}^n)$. If $x, y \in \mathbf{R}^n$, we set $x \cdot y = \sum_1^n x_j y_j$.

It will be convenient to have a compact notation for partial derivatives. We shall write ∂_j for $\partial/\partial x_j$, and for higher-order derivatives we use multi-index notation. A **multi-index** is an ordered n-tuple of nonnegative integers. If

$\alpha = (\alpha_1, \ldots, \alpha_n)$ is a multi-index, we set

$$|\alpha| = \sum_1^n \alpha_j, \quad \alpha! = \prod_1^n \alpha_j!, \quad \partial^\alpha = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n},$$

and for $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$,

$$x^\alpha = \prod_1^n x_j^{\alpha_j}.$$

Thus, for example, Taylor's formula for functions $f \in C^k$ reads

$$f(x) = \sum_{|\alpha| \leq k} (\partial^\alpha f)(x_0) \frac{(x - x_0)^\alpha}{\alpha!} + R_k(x)$$

$$\text{where } \lim_{x \to x_0} \frac{|R_k(x)|}{|x - x_0|^k} = 0,$$

and the product rule for derivatives becomes

$$\partial^\alpha(fg) = \sum_{\beta+\gamma=\alpha} \frac{\alpha!}{\beta!\gamma!} (\partial^\beta f)(\partial^\gamma g)$$

(Exercise 1).

We shall often avail ourselves of the sloppy but handy device of using the same notation for a function and its value at a point. Thus, "x^α" may be used to denote the function whose value at any x is x^α.

Two spaces of C^∞ functions on $\mathbf{R}^n$ will be of particular importance for us. The first is the **Schwartz space** $\mathcal{S}$ consisting of those C^∞ functions which, together with all their derivatives, vanish at infinity faster than any power of $|x|$. More precisely, for any nonnegative integer N and any multi-index α we define

$$\|f\|_{(N,\alpha)} = \sup_{x \in \mathbf{R}^n} (1 + |x|)^N |\partial^\alpha f(x)|$$

and

$$\mathcal{S} = \{ f \in C^\infty : \|f\|_{(N,\alpha)} < \infty \text{ for all } N, \alpha \}.$$

Examples of functions in $\mathcal{S}$ are easy to find: for instance, $f_\alpha(x) = x^\alpha e^{-|x|^2}$ where α is any multi-index.

(8.1) Proposition. $\mathcal{S}$ is a Fréchet space with the topology defined by the norms $\| \ \|_{(N,\alpha)}$.

Proof. The only nontrivial point is completeness. If $\{f_k\}$ is a Cauchy sequence in $\mathscr{S}$, then $\|f_j - f_k\|_{(N,\alpha)} \to 0$ for all N, α. In particular, $\{\partial^\alpha f_k\}$ converges uniformly for each α to a function g_α. Denoting by v_j the jth standard basis vector $(0,\dots,1,\dots,0)$ for $\mathbf{R}^n$, we have

$$f_k(x + tv_j) - f_k(x) = \int_0^t \partial_j f_k(x + sv_j)\, ds.$$

Letting $k \to \infty$ we obtain

$$g_0(x + tv_j) - g_0(x) = \int_0^t g_{v_j}(x + sv_j)\, ds.$$

The fundamental theorem of calculus implies that $g_{v_j} = \partial_j g_0$, and an induction on $|\alpha|$ then yields $g_\alpha = \partial^\alpha g_0$ for all α. It is then easy to check that $\|f_k - g_0\|_{(N,\alpha)} \to 0$ for all N, α. □

Another useful characterization of $\mathscr{S}$ is the following:

(8.2) Proposition. If $f \in C^\infty$, then $f \in \mathscr{S}$ iff $x^\alpha \partial^\beta f$ is bounded for all multi-indices α, β iff $\partial^\alpha(x^\beta f)$ is bounded for all multi-indices α, β.

Proof. Obviously $|x^\beta| \le (1 + |x|)^N$ for $|\beta| \le N$, while if δ is the (positive) minimum value of $\sum_1^n |x_j^N|$ on the unit sphere $|x| = 1$,

$$(1 + |x|)^N \le 2^N(1 + |x|^N) \le 2^N\left(1 + \delta^{-1}\sum_1^n |x_j^N|\right),$$

which implies the first equivalence. The second one follows from the fact that each $\partial^\alpha(x^\beta f)$ is a linear combination of terms of the form $x^\gamma \partial^\delta f$ and vice versa, by the product rule. □

The second space of C^∞ functions we shall need is the class C_c^∞ of C^∞ functions with compact support. It is obvious that $C_c^\infty \subset \mathscr{S}$ but a little less obvious that $C_c^\infty \ne \{0\}$. The standard construction of functions in C_c^∞ is based on the fact that the function f defined on $\mathbf{R}$ by $f(t) = e^{-1/t}$ for $t > 0$ and $f(t) = 0$ for $t \le 0$ is C^∞ even at the origin (Exercise 3). If we set $\psi(x) = f(1 - |x|^2)$, we have $\psi > 0$ on the open unit ball B and $\psi \in C_c^\infty(\bar{B})$. In the next section we shall use this single function to manufacture elements of C_c^∞ in great profusion; see Propositions (8.17) and (8.18).

We next investigate the continuity of translations on various function spaces. If f is any function on $\mathbf{R}^n$ and $y \in \mathbf{R}^n$, we define

$$f^y(x) = f(x - y). \tag{8.3}$$

We observe that $\|f^y\|_p = \|f\|_p$ for $1 \le p \le \infty$ and that $\|f^y\|_u = \|f\|_u$. f is called **uniformly continuous** if $\|f^y - f\|_u \to 0$ as $y \to 0$. (The reader should pause to check that this is equivalent to the usual ε-δ definition of uniform continuity.)

(8.4) Lemma. If $f \in C_c(\mathbf{R}^n)$, then f is uniformly continuous.

Proof. If $\varepsilon > 0$, for each $x \in \operatorname{supp}(f)$ there exists $\delta_x > 0$ such that $|f(x - y) - f(x)| < \varepsilon/2$ if $|y| < \delta_x$. Since $\operatorname{supp}(f)$ is compact, there exist $x_1, \dots, x_N$ such that the balls of radius δ_{x_j} about $x_j (1 \le j \le n)$ cover $\operatorname{supp}(f)$. If $\delta = \frac{1}{2}\min\{\delta_{x_j}\}$, then, one easily sees that $\|f^y - f\|_u < \varepsilon$ whenever $|y| < \delta$. □

(8.5) Proposition. If $1 \le p < \infty$, translation is continuous in the L^p norm: that is, if $f \in L^p$, $\lim_{y\to 0}\|f^y - f\|_p = 0$.

Proof. If $g \in C_c$, for $|y| < 1$ the functions g^y are all supported in a common compact set K, so by Lemma (8.4),

$$\int |g^y - g|^p \le \|g^y - g\|_u^p m(K) \to 0 \text{ as } y \to 0.$$

If $f \in L^p$ and $\varepsilon > 0$, by Proposition (7.9) there exists $g \in C_c$ with $\|g - f\|_u < \varepsilon/3$, whence

$$\|f^y - f\|_p \le \|(f - g)^y\|_p + \|g^y - g\|_p + \|g - f\|_p < \frac{2\varepsilon}{3} + \|g^y - g\|_p,$$

and $\|g^y - g\|_p < \varepsilon/3$ if y is sufficiently small. □

Proposition (8.4) is false for $p = \infty$, as one should expect since the L^∞ norm is closely related to the uniform norm: see Exercise 4.

Some of our results will be concerned with multiply periodic functions in $\mathbf{R}^n$, and for simplicity we take the fundamental period in each variable to be equal to 1. That is, we define a function f on $\mathbf{R}^n$ to be **periodic** if $f(x + k) = f(x)$ for all $k \in \mathbf{Z}^n$. Every periodic function is thus completely determined by its values on the unit cube $Q = [-\frac{1}{2}, \frac{1}{2})^n$. Periodic functions may be regarded as functions in the space $\mathbf{R}^n/\mathbf{Z}^n$ of cosets of $\mathbf{Z}^n$, which we call the ***n*-dimensional torus** and denote by $\mathbf{T}^n$. $\mathbf{T}^n$ is a compact Hausdorff space; it may be identified with the set of all $z = (z_1, \dots, z_n) \in \mathbf{C}^n$ such that $|z_j| = 1$ for all j, via the map

$$(x_1, \dots, x_n) \to (e^{2\pi i x_1}, \dots, e^{2\pi i x_n}).$$

On the other hand, for measure-theoretic purposes we identify $\mathbf{T}^n$ with the unit cube Q, and when we speak of Lebesgue measure on $\mathbf{T}^n$ we mean the measure

induced on $\mathbf{T}^n$ by Lebesgue measure on Q. Functions on $\mathbf{T}^n$ may be considered as periodic functions on $\mathbf{R}^n$ or as functions on Q; the point of view will be clear from the context when it matters.

EXERCISES

1. Prove the versions of Taylor's formula and the product rule stated in the text.
2. Observe that the binomial theorem can be written as follows:

$$(x_1 + x_2)^k = \sum_{|\alpha|=k} \frac{k!}{\alpha!} x^\alpha \quad [x = (x_1, x_2)].$$

Prove the following generalizations:

(a) The multinomial theorem: if $x \in \mathbf{R}^n$,

$$(x_1 + \cdots + x_n)^k = \sum_{|\alpha|=k} \frac{k!}{\alpha!} x^\alpha.$$

(b) The n-dimensional binomial theorem: if $x, y \in \mathbf{R}^n$,

$$(x + y)^\alpha = \sum_{\beta+\gamma=\alpha} \frac{\alpha!}{\beta!\gamma!} x^\beta y^\gamma.$$

3. Let $f(t) = e^{-1/t}$ for $t > 0$, $f(t) = 0$ for $t \leq 0$.
   (a) For $k \in \mathbf{N}$ and $t > 0$, $f^{(k)}(t) = P_k(1/t)e^{-1/t}$ where P_k is a polynomial.
   (b) $f^{(k)}(0)$ exists and equals zero for all $k \in \mathbf{N}$.
4. If $f \in L^\infty$ and $\|f^y - f\| \to 0$ as $y \to 0$, then f agrees a.e. with a uniformly continuous function. [Let $A_r f$ be as in Theorem (3.18). Then $A_r f$ is uniformly continuous for $r > 0$ and uniformly Cauchy as $r \to 0$.]

2. CONVOLUTIONS

Let f and g be measurable functions on $\mathbf{R}^n$. The **convolution** of f and g is the function $f * g$ defined by

$$f * g(x) = \int f(x - y)g(y)\, dy$$

for all x such that the integral exists. [In what follows we will need the fact

that $K(x, y) = f(x - y)$ is measurable on $\mathbf{R}^n \times \mathbf{R}^n$. We have $K = f \circ s$ where $s(x, y) = x - y$, so K is clearly Borel measurable if f is Borel measurable. This can always be assumed without affecting the definition of $f * g$, by Proposition (2.12). However, the Lebesgue measurability of K also follows from the Lebesgue measurability of f: see Exercise 5.] The elementary properties of convolutions are summarized in the following proposition.

(8.6) Proposition. Assuming that all integrals in question exist, we have:

(a) $f * g = g * f$.
(b) $(f * g) * h = f * (g * h)$.
(c) $(f * g)^z = f^z * g = f * g^z$ [notation as in (8.3)].
(d) $\operatorname{supp}(f * g) \subset \operatorname{supp}(f) + \operatorname{supp}(g)$ [$= \{x + y: x \in \operatorname{supp}(f), y \in \operatorname{supp}(g)\}$].

Proof. (a) is proved by the substitution $z = x - y$:

$$f * g(x) = \int f(x - y)g(y)\, dy = \int f(z)g(x - z)\, dz = g * f(x).$$

(b) follows from (a) and Fubini's theorem:

$$(f * g) * h(x) = \iint f(y)g(x - z - y)h(z)\, dy\, dz$$

$$= \iint f(y)g(x - y - z)h(z)\, dz\, dy = f * (g * h)(x).$$

As for (c),

$$(f * g)^z(x) = \int f(x - z - y)g(y)\, dy = \int f^z(x - y)g(y)\, dy = f^z * g(x),$$

and by (a),

$$(f * g)^z = (g * f)^z = g^z * f = f * g^z.$$

For (d), we observe that if $x \notin \operatorname{supp}(f) + \operatorname{supp}(g)$, then for any $y \in \operatorname{supp}(g)$ we have $x - y \notin \operatorname{supp}(f)$; hence $f(x - y)g(y) = 0$ for all y, so $f * g(x) = 0$. □

Various conditions can be imposed on f and g to guarantee that $f * g$ exists, at least almost everywhere. For example, if f is bounded and compactly supported, g can be any locally integrable function. The following two propositions contain the basic facts about convolutions of L^p functions.

(8.7) Young's Inequality. If $f \in L^1$ and $g \in L^p$ $(1 \le p \le \infty)$, then $f * g(x)$ exists for almost every x, $f * g \in L^p$, and $\|f * g\|_p \le \|f\|_1 \|g\|_p$.

Proof. This is a special case of Theorem (6.18), with $K(x, y) = f(x - y)$. Alternatively, one can use Minkowski's inequality for integrals:

$$\|f * g\|_p = \left\|\int f(y) g(\cdot - y)\, dy\right\|_p \le \int |f(y)| \|g^y\|_p dy = \|f\|_1 \|g\|_p. \qquad \square$$

(8.8) Proposition. If p and q are conjugate exponents, $f \in L^p$, and $g \in L^q$, then $f * g(x)$ exists for every x, $f * g$ is bounded and uniformly continuous, and $\|f * g\|_\infty \le \|f\|_p \|g\|_q$. If $1 < p < \infty$ (so that $1 < q < \infty$ also), then $f * g \in C_0(\mathbf{R}^n)$.

Proof. The existence of $f * g$ and the estimate for $\|f * g\|_\infty$ follow immediately from Hölder's inequality. In view of Propositions (8.5) and (8.6), so does the uniform continuity: if $1 \le p < \infty$,

$$\|(f * g)^y - f * g\|_\infty = \|(f^y - f) * g\|_\infty \le \|f^y - f\|_p \|g\|_q \to 0 \text{ as } y \to 0.$$

(If $p = \infty$, interchange the roles of f and g.) Finally, if $1 < p, q < \infty$, choose sequences $\{f_n\}$, $\{g_n\}$ of functions with compact support such that $\|f_n - f\|_p \to 0$ and $\|g_n - g\|_q \to 0$. By Proposition (8.6d) and what we have just proved, $f_n * g_n \in C_c$. But $f_n * g_n \to f * g$ uniformly, so $f * g \in C_0$. $\square$

The preceding results are all we shall use, but for the sake of completeness we state also the following generalization:

(8.9) Theorem. Suppose $1 \le p, q, r \le \infty$ and $p^{-1} + q^{-1} = r^{-1} + 1$.

(a) (General form of Young's inequality) If $f \in L^p$ and $g \in L^q$, then $f * g \in L^r$ and $\|f * g\|_r \le \|f\|_p \|g\|_q$.

(b) Suppose also that $p > 1$, $q > 1$, and $r < \infty$. If $f \in L^p$ and $g \in \text{weak}\, L^q$, then $f * g \in L^r$ and $\|f * g\|_r \le C_{pq} \|f\|_p [g]_q$ where C_{pq} is independent of f and g.

(c) Suppose that $p = 1$ and $r = q > 1$. If $f \in L^1$ and $g \in \text{weak}\, L^q$, then $f * g \in \text{weak}\, L^q$ and $[f * g]_q \le C_q \|f\|_1 [g]_q$ where C_q is independent of f and g.

Proof. To prove (a), let q be fixed. The special cases $p = 1$, $r = q$ and $p = q/(q - 1)$, $r = \infty$ are Propositions (8.7) and (8.8). The general case then follows from the Riesz–Thorin interpolation theorem. (See also Exercise 6 for a direct proof.) (b) and (c) are special cases of Theorem (6.35). $\square$

One of the most important properties of convolutions is that, roughly speaking, $f * g$ is at least as smooth as either f or g, because formally we have

$$\partial^\alpha (f * g)(x) = \partial^\alpha \int f(x - y) g(y)\, dy$$

$$= \int \partial^\alpha f(x - y) g(y)\, dy = (\partial^\alpha f) * g(x),$$

and similarly $\partial^\alpha (f * g) = f * \partial^\alpha g$. To make this precise one needs only to impose conditions on f and g so that differentiation under the integral sign is legitimate. One such result is the following; see also Exercises 7 and 8.

(8.10) Proposition. If $f \in L^1$, $g \in C^k$, and $\partial^\alpha g$ is bounded for $|\alpha| \le k$, then $f * g \in C^k$ and $\partial^\alpha (f * g) = f * \partial^\alpha g$ for $|\alpha| \le k$.

Proof. This is clear from Theorem (2.27). $\square$

(8.11) Proposition. If $f, g \in \mathscr{S}$, then $f * g \in \mathscr{S}$.

Proof. $f * g \in C^\infty$ by Proposition (8.10). Since

$$1 + |x| \le 1 + |x - y| + |y| \le (1 + |x - y|)(1 + |y|), \tag{8.12}$$

we have

$$(1 + |x|)^N |\partial^\alpha (f * g)(x)| \le \int (1 + |x - y|)^N |\partial^\alpha f(x - y)| (1 + |y|)^N |g(y)|\, dy$$

$$\le \|f\|_{(N,\alpha)} \|g\|_{(N+n+1,\alpha)} \int (1 + |y|)^{-n-1}\, dy,$$

which is finite by Corollary (2.52). $\square$

Convolutions of functions on the torus $\mathbf{T}^n$ are defined just as for functions on $\mathbf{R}^n$. (If one regards functions on $\mathbf{T}^n$ as periodic functions on $\mathbf{R}^n$, of course, the integration is to be extended over the unit cube rather than $\mathbf{R}^n$.) All of the preceding results remain valid, with the same proofs.

The following theorem underlies many of the important applications of convolutions on $\mathbf{R}^n$. We introduce a bit of notation which will be used frequently hereafter: if φ is any function on $\mathbf{R}^n$ and $t > 0$, we set

$$\varphi_t(x) = t^{-n} \varphi\left(\frac{x}{t}\right). \tag{8.13}$$

We observe that if $\varphi \in L^1$, then $\int \varphi_t$ is independent of t, by Theorem (2.43):

$$\int \varphi_t = \int t^{-n} \varphi(x/t)\, dx = \int \varphi(y)\, dy = \int \varphi.$$

Moreover, the "mass" of φ_t becomes concentrated at the origin as $t \to 0$. (Draw a picture if this isn't clear.)

(8.14) Theorem. Suppose $\varphi \in L^1$ and $\int \varphi(x)\,dx = a$.

(a) If $f \in L^p$ $(1 \le p \le \infty)$, then $f * \varphi_t \to af$ in the L^p norm as $t \to 0$.

(b) If f is bounded and uniformly continuous, then $f * \varphi_t \to af$ uniformly as $t \to 0$.

(c) If $f \in L^\infty$ and f is continuous on an open set U, then $f * \varphi_t \to af$ uniformly on compact subsets of U as $t \to 0$.

Proof. We shall prove (a) and (b) simultaneously. Setting $y = tz$, we have

$$f * \varphi_t(x) - af(x) = \int [f(x-y) - f(x)]\varphi_t(y)\,dy$$

$$= \int [f(x - tz) - f(x)]\varphi(z)\,dz$$

$$= \int [f^{tz}(x) - f(x)]\varphi(z)\,dz.$$

Now we apply Minkowski's inequality for integrals:

$$\|f * \varphi_t - af\|_p \le \int \|f^{tz} - af\|_p |\varphi(z)|\,dz.$$

But $\|f^{tz} - af\|_p$ is bounded by $2\|f\|_p$ and tends to zero as $t \to 0$ for each z [by Proposition (8.5) in case (a) or by the uniform continuity of f in case (b), where we take $p = \infty$]. The desired result therefore follows from the dominated convergence theorem. As for (c), given $\varepsilon > 0$ let us choose a compact $E \subset \mathbf{R}^n$ such that $\int_{E^c} |\varphi| < \varepsilon$. Also, let K be a compact subset of U. If $x \in K$ and $z \in E$, then, we will have $x - tz \in U$ if t is sufficiently small, so from the compactness of K it follows as in Lemma (8.4) that

$$\sup_{x \in K, z \in E} |f(x - tz) - f(x)| < \varepsilon$$

for small t. But then

$$\sup_{x \in K} |f * \varphi_t(x) - af(x)|$$

$$\le \sup_{x \in K} \left[\int_E + \int_{E^c}\right] |f(x - tz) - f(x)||\varphi(z)|\,dz$$

$$\le \varepsilon \int |\varphi| + 2\|f\|_\infty \varepsilon,$$

from which (c) follows. □

If we impose slightly stronger conditions on φ, we can also show that $f * \varphi_t \to af$ almost everywhere for $f \in L^p$. The device in the following proof of breaking up an integral into pieces and estimating each piece separately is a standard trick of the trade in Fourier analysis.

(8.15) Theorem. Suppose $|\varphi(x)| \le C(1 + |x|)^{-n-\varepsilon}$ for some C, $\varepsilon > 0$ [which implies that $\varphi \in L^1$ by Corollary (2.52)], and $\int \varphi(x)\,dx = a$. If $f \in L^p$ $(1 \le p \le \infty)$, then $f * \varphi_t(x) \to af(x)$ for every x in the Lebesgue set of f—in particular, for almost every x, and for every x at which f is continuous.

Proof. If x is in the Lebesgue set of f, for any $\delta > 0$ there exists $\eta > 0$ such that

$$\int_{|y|<r} |f(x-y) - f(x)|\,dy \le \delta r^n \text{ for } r \le \eta. \tag{8.16}$$

Let us set

$$I_1 = \int_{|y| \le \eta} |f(x-y) - f(x)||\varphi_t(y)|\,dy,$$

$$I_2 = \int_{|y| > \eta} |f(x-y) - f(x)||\varphi_t(y)|\,dy.$$

We claim that I_1 is bounded by $A\delta$ where A is independent of t, while $I_2 \to 0$ as $t \to 0$. Since

$$|f * \varphi_t(x) - af(x)| \le I_1 + I_2,$$

we will have

$$\limsup_{t \to 0} |f * \varphi_t(x) - af(x)| \le A\delta,$$

and since δ is arbitrary, this will complete the proof.

To estimate I_1, let K be the integer such that $2^K \le \eta/t < 2^{K+1}$ if $\eta/t \ge 1$, and $K = 0$ if $\eta/t < 1$. We view the ball $|y| \le \eta$ as the union of the annuli $2^{-k-1}\eta < |y| \le 2^{-k}\eta$ $(0 \le k < K)$ and the ball $|y| \le 2^{-K}\eta$. On the kth annulus we use the estimate

$$|\varphi_t(y)| \le Ct^{-n}\left(\frac{2^{-k}\eta}{t}\right)^{-n-\varepsilon},$$

and on the ball $|y| \le 2^{-K}\eta$ we use the estimate $|\varphi_t(y)| \le Ct^{-n}$. Thus

$$I_1 \le \sum_0^{K-1} Ct^{-n}\left(\frac{2^{-k}\eta}{t}\right)^{-n-\varepsilon} \int_{2^{-k-1}\eta \le |y| \le 2^{-k}\eta} |f(x-y) - f(x)|\,dy$$

$$+ Ct^{-n} \int_{|y| \le 2^{-K}\eta} |f(x-y) - f(x)|\,dy.$$

Therefore, by (8.16) and the fact that $\eta/t \geq 2^{-K}$,

$$I_1 \leq C\delta \sum_0^{K-1} t^{-n}\left(\frac{2^{-k}\eta}{t}\right)^{-n-\varepsilon}(2^{-k}\eta)^n + C\delta t^{-n}(2^{-K}\eta)^n$$

$$= C\delta\left(\frac{\eta}{t}\right)^{-\varepsilon}\sum_0^{K-1} 2^{k\varepsilon} + C\delta\left(\frac{2^{-K}\eta}{t}\right)^n$$

$$= C\delta\left(\frac{\eta}{t}\right)^{-\varepsilon}\frac{2^{K\varepsilon}-1}{2^{\varepsilon}-1} + C\delta\left(\frac{2^{-K}\eta}{t}\right)^n$$

$$\leq C\left[(2^{\varepsilon}-1)^{-1} + 2^n\right]\delta.$$

As for I_2, if p' is the conjugate exponent to p and χ is the characteristic function of $\{y\colon |y| > \eta\}$, by Hölder's inequality we have

$$I_2 \leq \int_{|y|>\eta} [|f(x-y)| + |f(x)|]\,|\varphi_t(y)|\,dy$$

$$\leq \|f\|_p \|\chi\varphi_t\|_{p'} + |f(x)|\,\|\chi\varphi_t\|_1,$$

so it suffices to show that for $1 \leq q \leq \infty$, and in particular for $q = 1$ and $q = p'$, $\|\chi\varphi_t\|_q \to 0$ as $t \to 0$. If $q = \infty$, this is obvious:

$$\|\chi\varphi_t\|_\infty \leq Ct^{-n}[1 + (\eta/t)]^{-n-\varepsilon} = Ct^{\varepsilon}(t+\eta)^{-n-\varepsilon} \leq C\eta^{-n-\varepsilon}t^{\varepsilon}.$$

If $q < \infty$, by Corollary (2.51) we have

$$\|\chi\varphi_t\|_q^q = \int_{|y|>\eta} t^{-nq}|\varphi(y/t)|^q\,dy = t^{n(1-q)}\int_{|z|>\eta/t}|\varphi(z)|^q\,dz$$

$$= C_1 t^{n(1-q)}\int_{\eta/t}^{\infty} r^{n-1-(n+\varepsilon)q}\,dr = C_2 t^{n(1-q)}\left(\frac{\eta}{t}\right)^{n-(n+\varepsilon)q} = C_3 t^{\varepsilon q}.$$

In either case, $\|\chi\varphi_t\|_q$ is dominated by t^{ε}, so we are done. □

In most of the applications of the two preceding theorems one has $a = \int\varphi(x)dx = 1$, although the case $a = 0$ is also of interest. If $a = 1$, $\{\varphi_t\}_{t>0}$ is called an **approximate identity**, as it furnishes an approximation to the identity operator on L^p by convolution operators. This construction is useful for approximating L^p functions by functions having specified regularity properties. For example, we have the following important results:

(8.17) Proposition. If $1 \leq p < \infty$, C_c^∞ (and hence also $\mathcal{S}$) is dense in L^p.

Proof. If $f \in L^p$ and $\varepsilon > 0$, choose $g \in C_c$ with $\|f - g\|_p < \varepsilon/2$, by Proposition (7.9). Let φ be a function in C_c^∞ such that $\int\varphi = 1$—for example, take $\varphi = [\int\psi]^{-1}\psi$ where ψ is the function constructed in Section 1. Then $g * \varphi_t \in C_c^\infty$ by Propositions (8.6d) and (8.10), and $\|g * \varphi_t - g\|_p < \varepsilon/2$ for sufficiently small t by Theorem (8.14). □

(8.18) The C^∞ Urysohn Lemma. If $K \subset \mathbf{R}^n$ is compact and U is an open set containing K, there exists $f \in C_c^\infty$ such that $0 \leq f \leq 1$, $f = 1$ on K, and $\operatorname{supp}(f) \subset U$.

Proof. Let $\delta = \rho(K, U^c)$ (the distance from K to U^c, which is positive since K is compact), and let $V = \{x\colon \rho(x,K) < \delta/3\}$. Choose a nonnegative $\varphi \in C_c^\infty$ such that $\int\varphi = 1$ and $\varphi(x) = 0$ for $|x| \geq \delta/3$ (for example, a modification of the ψ in Section 1), and set $f = \chi_V * \varphi$. Then $f \in C_c^\infty$ by Propositions (8.6d) and (8.10), and it is easily checked that $0 \leq f \leq 1$, $f = 1$ on K, and

$$\operatorname{supp}(f) \subset \{x\colon \rho(x,K) \leq 2\delta/3\} \subset U.$$ □

EXERCISES

5. If $s\colon \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ is defined by $s(x-y) = x - y$, then $s^{-1}(E)$ is Lebesgue measurable whenever E is Lebesgue measurable. [For $n = 1$, draw a picture of $s^{-1}(E) \subset \mathbf{R}^2$. It should be clear that after rotation through an angle $\pi/4$, $s^{-1}(E)$ becomes $F \times \mathbf{R}$ where $F = \{x/\sqrt{2}\colon x \in E\}$, and Theorem (2.43) can be applied. The same idea works in higher dimensions.]

6. Prove Theorem (8.9a) by using Exercise 6.31 to show that

$$|f * g(x)|^r \leq \left(\int |f(y)|^p |g(x-y)|^q\,dy\right)\|f\|_p^{r-p}\|g\|_q^{r-q}.$$

7. If f is locally integrable on $\mathbf{R}^n$ and $g \in C^k$ has compact support, then $f * g \in C^k$.

8. Suppose $f \in L^p(\mathbf{R})$. If there exists $h \in L^p(\mathbf{R})$ such that

$$\lim_{y\to 0}\|y^{-1}(f^{-y} - f) - h\|_p = 0,$$

we call h the **(strong) L^p derivative** of f and write $h = df/dx$. If $f \in L^p(\mathbf{R}^n)$, L^p partial derivatives of f are defined similarly. If p and q are conjugate exponents, $f \in L^p$, $g \in L^q$, and the L^p derivative $\partial_j f$ exists, then $\partial_j(f * g)$ exists (in the ordinary sense) and equals $(\partial_j f) * g$.

9. If $f \in L^p(\mathbf{R})$, the L^p derivative of f (call it h) exists iff f is absolutely continuous on every bounded interval (after modification on a null set) and its pointwise derivative f' is in L^p, in which case $h = f'$ a.e. [For "only if," use Exercise 8: take $g \in C_c$ with $\int g = 1$, so that $f * g_t \to f$ and $(f * g_t)' \to h$ as $t \to 0$. For "if," write

$$y^{-1}[f(x+y) - f(x)] - f'(x) = y^{-1}\int_0^y [f'(x+t) - f'(x)]\,dt$$

and use Minkowski's inequality for integrals.]

10. Let φ satisfy the hypotheses of Theorem (8.15). If $f \in L^p$ $(1 \le p \le \infty)$, define the **φ-maximal function** of f to be $M_\varphi f(x) = \sup_{t>0} |f * \varphi_t(x)|$. (Observe that the Hardy–Littlewood maximal function Hf is $M_\varphi |f|$ where φ is the characteristic function of the unit ball divided by the volume of the ball.) Show that there is a constant C, independent of f, such that $M_\varphi f \le CHf$. [Break up the integral $\int f(x-y)\varphi_t(y)\,dy$ as the sum of the integrals over $|y| \le t$ and over $2^k t < |y| \le 2^{k+1}t$ $(k = 0, 1, 2, \dots)$, and estimate φ_t on each region.] It follows from Theorem (3.17) that M_φ is weak type $(1,1)$, and the proof of Theorem (3.18) can then be adapted to give an alternate demonstration that $f * \varphi_t \to [\int \varphi] f$ a.e.

11. Young's inequality shows that L^1 is a Banach algebra with convolution as multiplication.

(a) If $\mathcal{I}$ is an ideal in the algebra L^1, so is its closure in L^1.

(b) If $f \in L^1$, the smallest closed ideal in L^1 containing f is the smallest closed subspace of L^1 containing all translates of f. [If $g \in C_c$, $f * g(x)$ can be approximated by sums $\Sigma f(x - y_j)g(y_j)$. On the other hand, if $\{\varphi_t\}$ is an approximate identity, $f * (\varphi_t)^y \to f^y$ as $t \to 0$.]

3. THE FOURIER TRANSFORM

One of the fundamental principles of harmonic analysis is the exploitation of symmetry. To be more specific, if one is doing analysis on a space on which a group acts, it is a good idea to study functions (or other analytic objects) which transform in simple ways under the group action, and then try to decompose arbitrary functions as sums or integrals of these basic functions.

The spaces we are studying are $\mathbf{R}^n$ and $\mathbf{T}^n$, which are Abelian groups under addition and act on themselves by translation. The building blocks of harmonic analysis on these spaces are the functions which transform under translation by multiplication by a factor of absolute value one, that is, functions f such that for each x there is a number $\varphi(x)$ with $|\varphi(x)| = 1$ such that $f(y + x) = \varphi(x)f(y)$. We observe that if f and φ have this property, then $f(x) = \varphi(x)f(0)$, so that f is completely determined by φ once $f(0)$ is given; moreover, $\varphi(x + y)f(0) = f(x + y) = \varphi(x)\varphi(y)f(0)$, so that (unless $f = 0$), $\varphi(x + y) = \varphi(x)\varphi(y)$. In short, to find all f's which transform as described above, it suffices to find all φ's of absolute value one which satisfy the functional equation $\varphi(x + y) = \varphi(x)\varphi(y)$. Upon imposing the natural requirement that φ should be measurable, we have a complete solution to this problem.

(8.19) Theorem. If φ is a measurable function on $\mathbf{R}^n$ (resp. $\mathbf{T}^n$) such that $\varphi(x + y) = \varphi(x)\varphi(y)$ and $|\varphi| = 1$, there exists $\xi \in \mathbf{R}^n$ (resp. $\xi \in \mathbf{Z}^n$) such that $\varphi(x) = e^{2\pi i x \cdot \xi}$.

Proof. We first prove this assertion for $\mathbf{R}^1$. Let $a \in \mathbf{R}$ be such that $\int_0^a \varphi(t)\,dt \neq 0$; such an a surely exists, for otherwise the Lebesgue differentiation theorem would imply that $\varphi = 0$ a.e. Setting $A = [\int_0^a \varphi(t)\,dt]^{-1}$, then, we have

$$\varphi(x) = A\int_0^a \varphi(x)\varphi(t)\,dt = A\int_0^a \varphi(x + t)\,dt = A\int_x^{x+a} \varphi(t)\,dt.$$

Thus φ, being the indefinite integral of a locally integrable function, is continuous; and then, being the integral of a continuous function, it is C^1. Moreover,

$$\varphi'(x) = A[\varphi(x + a) - \varphi(x)] = B\varphi(x) \text{ where } B = A[\varphi(a) - 1].$$

It follows that $(d/dx)(e^{-Bx}\varphi(x)) = 0$, so that $e^{-Bx}\varphi(x)$ is constant. Since $\varphi(0) = 1$, we have $\varphi(x) = e^{Bx}$, and since $|\varphi| = 1$, B is purely imaginary, and we can write $B = 2\pi i\xi$ for some $\xi \in \mathbf{R}$. This completes the proof for $\mathbf{R}^1$; as for $\mathbf{T}^1$, the φ we have been considering will be periodic (with period 1) iff $e^{2\pi i\xi} = 1$ iff $\xi \in \mathbf{Z}$.

The n-dimensional case follows easily, for if $v_1, \dots, v_n$ is the standard basis for $\mathbf{R}^n$, the functions $\psi_j(t) = \varphi(tv_j)$ satisfy $\psi_j(t + s) = \psi_j(t)\psi_j(s)$ on $\mathbf{R}$, so that $\psi_j(t) = e^{2\pi i t\xi_j}$ and hence

$$\varphi(x) = \varphi\left(\sum_1^n x_j v_j\right) = \prod_1^n \psi_j(x_j) = e^{2\pi i x \cdot \xi}. \qquad \square$$

The idea now is to decompose more or less arbitrary functions on $\mathbf{T}^n$ or $\mathbf{R}^n$ as sums or integrals of the exponentials $e^{2\pi i x \cdot \xi}$. In the case of $\mathbf{T}^n$ this works out very simply, in the following sense.

(8.20) Theorem. Let $E_\kappa(x) = e^{2\pi i \kappa \cdot x}$. Then $\{E_\kappa : \kappa \in \mathbf{Z}^n\}$ is an orthonormal basis for $L^2(\mathbf{T}^n)$.

Proof. Verification of orthonormality is an easy exercise in calculus: by Fubini's theorem it boils down to the fact that $\int_0^1 e^{2\pi i k t}\,dt$ equals 1 if $k = 0$ and

equals 0 otherwise. Next, since $E_\kappa E_\lambda = E_{\kappa+\lambda}$, the set of finite linear combinations of the E_κ's is an algebra, and it clearly separates points. Also, $E_0 = 1$ and $\overline{E}_\kappa = E_{-\kappa}$. Since $\mathbf{T}^n$ is compact, the Stone–Weierstrass theorem implies that this algebra is dense in $C(\mathbf{T}^n)$ in the uniform norm and hence in the L^2 norm, and $C(\mathbf{T}^n)$ is itself dense in $L^2(\mathbf{T}^n)$ by Proposition (7.9). It follows that $\{E_\kappa\}$ is a basis. □

To restate this result: if $f \in L^2(\mathbf{T}^n)$, we define its **Fourier transform** $\hat{f} \in l^2(\mathbf{Z}^n)$ by

$$\hat{f}(\kappa) = \langle f, E_\kappa \rangle = \int_{\mathbf{T}^n} f(x) e^{-2\pi i \kappa \cdot x}\, dx.$$

Then $f = \sum_{\kappa \in \mathbf{Z}^n} \hat{f}(\kappa) E_\kappa$, where the series converges in $L^2(\mathbf{T}^n)$. Moreover, if $\sum_{\kappa \in \mathbf{Z}^n} |\hat{f}(\kappa)| < \infty$, the series $\sum_{\kappa \in \mathbf{Z}^n} \hat{f}(\kappa) E_\kappa$ converges absolutely and uniformly, and its sum is f a.e. (Cf. Exercise 6.9.) Actually, the definition of $\hat{f}(\kappa)$ makes sense if f is merely in $L^1(\mathbf{T}^n)$, and $|\hat{f}(\kappa)| \le \|f\|_1$, so the Fourier transform extends to a norm-decreasing map from $L^1(\mathbf{T}^n)$ to $l^\infty(\mathbf{Z}^n)$. (In the next section we shall consider the problem of recovering f from $\hat{f}$ when f is merely in L^1.) More generally, we have:

(8.21) The Hausdorff–Young Inequality. Suppose p and q are conjugate exponents and $1 \le p \le 2$. If $f \in L^p(\mathbf{T}^n)$, then $\hat{f} \in l^q(\mathbf{Z}^n)$, and $\|\hat{f}\|_q \le \|f\|_p$.

Proof. Since $\|\hat{f}\|_\infty \le \|f\|_1$ and $\|\hat{f}\|_2 = \|f\|_2$ for $f \in L^1$ or $f \in L^2$, the assertion follows from the Riesz–Thorin interpolation theorem. □

The situation on $\mathbf{R}^n$ is more delicate. The formal analog of Theorem (8.20) should be

$$\hat{f}(\xi) = \int_{\mathbf{R}^n} f(x) e^{-2\pi i x \cdot \xi}\, dx, \qquad f(x) = \int_{\mathbf{R}^n} \hat{f}(\xi) e^{2\pi i x \cdot \xi}\, d\xi.$$

These relations turn out to be valid when suitably interpreted, but some care is needed. In the first place, the integral defining $\hat{f}(\xi)$ is likely to diverge when $f \in L^2$. However, it certainly converges if $f \in L^1$. We therefore begin by defining the **Fourier transform** of $f \in L^1(\mathbf{R}^n)$ to be

$$\mathcal{F}f(\xi) = \hat{f}(\xi) = \int_{\mathbf{R}^n} f(x) e^{-2\pi i x \cdot \xi}\, dx.$$

(We use the notation $\mathcal{F}$ for the Fourier transform only in certain situations where it is needed for clarity.) Clearly $\|\hat{f}\|_\infty \le \|f\|_1$, and $\hat{f}$ is continuous by Theorem (2.27); thus

$$\mathcal{F}\colon L^1(\mathbf{R}^n) \to BC(\mathbf{R}^n).$$

We summarize the elementary properties of $\mathcal{F}$ in a theorem.

(8.22) Theorem. Suppose $f, g \in L^1(\mathbf{R}^n)$.

(a) With notation as in (8.3), we have $(f^y)\hat{}\,(\xi) = e^{-2\pi i y \cdot \xi} \hat{f}(\xi)$, and $(\hat{f})^\eta = \hat{h}$ where $h(x) = e^{2\pi i x \cdot \eta} f(x)$.

(b) If T is an invertible linear transformation of $\mathbf{R}^n$ and $S = (T^*)^{-1}$ is its inverse transpose, then $(f \circ T)\hat{} = |\det T|^{-1} \hat{f} \circ S$. In particular, if T is a rotation then $(f \circ T)\hat{} = \hat{f} \circ T$, and if $Tx = t^{-1}x$ $(t > 0)$ then $(f \circ T)\hat{}\,(\xi) = t^n \hat{f}(t\xi)$ [so that $(f_t)\hat{}\,(\xi) = \hat{f}(t\xi)$ in the notation of (8.13)].

(c) $(f * g)\hat{} = \hat{f}\hat{g}$.

(d) If $x^\alpha f \in L^1$ for $|\alpha| \le k$, then $\hat{f} \in C^k$ and $\partial^\alpha \hat{f} = ((-2\pi i x)^\alpha f)\hat{}$.

(e) If $f \in C^k$, $\partial^\alpha f \in L^1$ for $|\alpha| \le k$, and $\partial^\alpha f \in C_0$ for $|\alpha| \le k - 1$, then $(\partial^\alpha f)\hat{}\,(\xi) = (2\pi i \xi)^\alpha \hat{f}(\xi)$.

(f) (The Riemann–Lebesgue Lemma) $\mathcal{F}(L^1(\mathbf{R}^n)) \subset C_0(\mathbf{R}^n)$.

Proof. (a) We have

$$(f^y)\hat{}\,(\xi) = \int f(x - y) e^{-2\pi i x \cdot \xi}\, dx = \int f(x) e^{-2\pi i (x+y) \cdot \xi}\, dx = e^{-2\pi i y \cdot \xi} \hat{f}(\xi),$$

and similarly for the other formula.

(b) By Theorem (2.43),

$$(f \circ T)\hat{}\,(\xi) = \int f(Tx) e^{-2\pi i x \cdot \xi}\, dx = |\det T|^{-1} \int f(x) e^{-2\pi i T^{-1}x \cdot \xi}\, dx$$

$$= |\det T|^{-1} \int f(x) e^{-2\pi i x \cdot S\xi}\, dx = |\det T|^{-1} \hat{f}(S\xi).$$

(c) By Fubini's theorem,

$$(f * g)\hat{}\,(\xi) = \iint f(x - y) g(y) e^{-2\pi i x \cdot \xi}\, dy\, dx$$

$$= \iint f(x - y) e^{-2\pi i (x-y) \cdot \xi} g(y) e^{-2\pi i y \cdot \xi}\, dx\, dy$$

$$= \hat{f}(\xi) \int g(y) e^{-2\pi i y \cdot \xi}\, dy$$

$$= \hat{f}(\xi) \hat{g}(\xi).$$

(d) By Theorem (2.27) and induction on $|\alpha|$,

$$\partial^\alpha \hat{f}(\xi) = \partial_\xi^\alpha \int f(x) e^{-2\pi i x \cdot \xi}\, dx = \int f(x) (-2\pi i x)^\alpha e^{-2\pi i x \cdot \xi}\, dx.$$

(e) First assume $n = |\alpha| = 1$. Since $f \in C_0$, we can integrate by parts:

$$\int f'(x)e^{-2\pi ix\cdot\xi}\,dx = f(x)e^{-2\pi ix\cdot\xi}\Big|_{-\infty}^{\infty} - \int f(x)(-2\pi i\xi)e^{-2\pi ix\cdot\xi}\,dx$$

$$= 2\pi i\xi\hat{f}(\xi).$$

The argument for $n > 1$, $|\alpha| = 1$ is the same—to compute $(\partial_j f)\hat{}$, integrate by parts in the jth variable—and the general case follows by induction on $|\alpha|$.

(f) By (e), if $f \in C^1 \cap C_c$, then $|\xi|\hat{f}(\xi)$ is bounded and hence $\hat{f} \in C_0$. But the set of all such f's is dense in L^1 by Proposition (8.17), and $\hat{f}_n \to \hat{f}$ uniformly whenever $f_n \to f$ in L^1. Since C_0 is complete in the uniform norm, the result follows. □

Parts (d) and (e) of Theorem (8.22) point to a fundamental property of the Fourier transform: smoothness properties of f are reflected in the rate of decay of $\hat{f}$ at infinity, and vice versa. Parts (a), (c), (e), and (f) of this theorem are valid also on $\mathbf{T}^n$, as is (b) provided that T leaves the lattice $\mathbf{Z}^n$ invariant: see Exercise 12.

(8.23) Corollary. $\mathscr{F}(\mathscr{S}) \subset \mathscr{S}$.

Proof. If $f \in \mathscr{S}$, then $x^\alpha\partial^\beta f \in L^1 \cap C_0$ for all α, β by Corollary (2.52), so by Theorem (8.22d, e), $\hat{f}$ is C^∞ and

$$(x^\alpha\partial^\beta f)\hat{} = (-i)^{|\alpha|}(2\pi i)^{|\alpha|-|\beta|}\partial^\alpha(\xi^\beta\hat{f}).$$

Thus $\partial^\alpha(\xi^\beta\hat{f})$ is bounded for all α, β, whence $\hat{f} \in \mathscr{S}$ by Proposition (8.2). □

At this point, let us compute an important specific Fourier transform.

(8.24) Proposition. If $a > 0$ and $f(x) = e^{-\pi a|x|^2}$, then $\hat{f}(\xi) = a^{-n/2}e^{-\pi|\xi|^2/a}$.

Proof. First consider the case $n = 1$. By Theorem (8.22d, e),

$$(\hat{f})'(\xi) = (-2\pi ixe^{-\pi ax^2})\hat{}(\xi) = \left(\frac{i}{a}\right)\left((e^{-\pi ax^2})'\right)\hat{}(\xi)$$

$$= \left(\frac{i}{a}\right)(2\pi i\xi)\hat{f}(\xi) = -\left(\frac{2\pi}{a}\right)\xi\hat{f}(\xi).$$

It follows that $(d/d\xi)(e^{\pi\xi^2/a}\hat{f}(\xi)) = 0$, so that $e^{\pi\xi^2/a}\hat{f}(\xi)$ is constant. To evaluate the constant, set $\xi = 0$ and use Proposition (2.53):

$$\hat{f}(0) = \int e^{-\pi ax^2}\,dx = a^{-1/2}.$$

The n-dimensional case follows from Fubini's theorem, since $|x|^2 = \sum_1^n x_j^2$:

$$\hat{f}(\xi) = \prod_1^n \int \exp\left(-\pi ax_j^2 - 2\pi ix_j\xi_j\right)dx_j$$

$$= \prod_1^n \left[a^{-1/2}\exp\left(\frac{-\pi\xi_j^2}{a}\right)\right] = a^{-n/2}\exp\left(\frac{-\pi|\xi|^2}{a}\right).$$ □

We are now ready to invert the Fourier transform. If $f \in L^1$, we define

$$\check{f}(x) = \hat{f}(-x) = \int f(\xi)e^{2\pi ix\cdot\xi}\,d\xi,$$

and we claim that if $f \in L^1$ and $\hat{f} \in L^1$ then $(\hat{f})\check{} = f$. A simple appeal to Fubini's theorem fails because the integrand in

$$(\hat{f})\check{}(x) = \iint f(y)e^{-2\pi iy\cdot\xi}e^{2\pi ix\cdot\xi}\,dy\,d\xi$$

is not in $L^1(\mathbf{R}^n \times \mathbf{R}^n)$. The trick is to introduce a convergence factor and then pass to the limit, using Fubini's theorem via the following lemma.

(8.25) Lemma. If $f, g \in L^1$, then $\int f\hat{g} = \int \hat{f}g$.

Proof. Both integrals are equal to $\iint f(x)g(\xi)e^{-2\pi ix\cdot\xi}dx\,d\xi$. □

(8.26) The Fourier Inversion Theorem. If $f \in L^1$ and $\hat{f} \in L^1$, then $(\hat{f})\check{} = (\check{f})\hat{} = f$ a.e.

Proof. Given $t > 0$ and $x \in \mathbf{R}^n$, set

$$\varphi(\xi) = \exp(2\pi ix\cdot\xi - \pi t^2|\xi|^2).$$

By Theorem (8.22a) and Proposition (8.24),

$$\hat{\varphi}(y) = t^{-n}\exp(-\pi|x-y|^2/t^2) = g_t(x-y) \text{ where } g(x) = e^{-\pi|x|^2}.$$

By Lemma (8.25), then,

$$\int e^{-\pi t^2|\xi|^2}e^{2\pi ix\cdot\xi}\hat{f}(\xi)\,d\xi = \int \hat{f}\varphi = \int f\hat{\varphi} = \int f(y)g_t(x-y)\,dy = f * g_t(x).$$

Since $\int e^{-\pi|x|^2} = 1$, by Theorem (8.14) we have $f * g_t \to f$ in the L^1 norm as $t \to 0$. On the other hand, since $\hat{f} \in L^1$ the dominated convergence theorem

yields

$$\int e^{-\pi t^2|\xi|^2} e^{2\pi i x\cdot\xi} \hat{f}(\xi)\,d\xi \to \int e^{2\pi i x\cdot\xi}\hat{f}(\xi)\,d\xi = (\hat{f})^{\vee}(x).$$

It follows that $f = (\hat{f})^{\vee}$ a.e., and similarly $(\check{f})\hat{} = f$ a.e. □

Remark. If $f \in L^1$ and $\hat{f} \in L^1$, then $(\hat{f})^{\vee} \in C_0$. Hence f can be made continuous by modification on a null set, and the equation $(\hat{f})^{\vee} = f$ then holds everywhere.

(8.27) Corollary. If $f \in L^1$ and $\hat{f} = 0$, then $f = 0$ a.e.

(8.28) Corollary. $\mathscr{F}$ is an isomorphism of $\mathscr{S}$ onto itself.

Proof. We know that if $f \in \mathscr{S}$, then $\hat{f} \in \mathscr{S}$ [Corollary (8.23)] and hence $\check{f} \in \mathscr{S}$ (by definition of $\check{f}$). Since $f = (\hat{f})^{\vee}$, the Fourier transform is bijective on $\mathscr{S}$, and continuity follows from the proof of Corollary (8.23). □

We are now ready to develop the L^2 theory of the Fourier transform.

(8.29) The Plancherel Theorem. If $f \in L^1 \cap L^2$, then $\hat{f} \in L^2$, and $\mathscr{F}|L^1 \cap L^2$ extends uniquely to a unitary isomorphism on L^2.

Proof. Let $\mathscr{X} = \{f \in L^1 : \hat{f} \in L^1\}$. Since $\hat{f} \in L^1$ implies $f \in L^\infty$, we have $\mathscr{X} \subset L^2$ by Proposition (6.10), and $\mathscr{X}$ is dense in L^2 because $\mathscr{S} \subset \mathscr{X}$ and $\mathscr{S}$ in dense in L^2. If $f, g \in \mathscr{X}$, let $h = \bar{\check{g}}$. By the inversion theorem,

$$\hat{h}(\xi) = \int e^{-2\pi i x\cdot\xi}\bar{\check{g}}(x)\,dx = \overline{\int e^{2\pi i x\cdot\xi}\check{g}(x)\,dx} = \overline{g(\xi)}\,.$$

Hence by Lemma (8.25),

$$\int f\bar{g} = \int \hat{f}\hat{h} = \int \hat{f}\bar{\hat{g}}.$$

Thus $\mathscr{F}|\mathscr{X}$ preserves the L^2 scalar product, and in particular, by taking $g = f$ we obtain $\|\hat{f}\|_2 = \|f\|_2$. Since $\mathscr{F}(\mathscr{X}) = \mathscr{X}$ by the inversion theorem, $\mathscr{F}|\mathscr{X}$ extends by continuity to a unitary isomorphism on L^2. It remains only to show that this extension agrees with $\mathscr{F}$ on $L^1 \cap L^2$. But if $f \in L^1 \cap L^2$ and $g(x) = e^{-\pi|x|^2}$ as in the proof of the inversion theorem, we have $f * g_t \in L^1 \cap L^2$ by Young's inequality and $(f * g_t)\hat{} \in L^1$ because $(f * g_t)\hat{}(\xi) = e^{-\pi t^2|\xi|^2}\hat{f}(\xi)$ and $\hat{f}$ is bounded. Hence $f * g_t \in \mathscr{X}$; moreover, by Theorem (8.14), $f * g_t \to f$ in both the L^1 and L^2 norms. Therefore $(f * g_t)\hat{} \to \hat{f}$ both uniformly and in the L^2 norm, and we are done. □

We have thus extended the domain of the Fourier transform from L^1 to $L^1 + L^2$. Just as on the torus, it follows from the Riesz–Thorin theorem that $\mathscr{F}$

maps L^p into L^q where $1 \le p \le 2$ and $p^{-1} + q^{-1} = 1$, and that the Hausdorff–Young inequality $\|\hat{f}\|_q \le \|f\|_p$ holds. What remains is the problem of recovering f from $\hat{f}$ in some reasonably explicit fashion when $\hat{f} \notin L^1$. The proof of the Fourier inversion theorem contains the essential idea: replace the divergent integral $\int \hat{f}(\xi)e^{2\pi i x\cdot\xi}\,d\xi$ by $\int \hat{f}(\xi)\Phi(t\xi)e^{2\pi i x\cdot\xi}\,d\xi$ where Φ is a continuous function which vanishes rapidly enough at infinity to make the integral converge. If we choose Φ to satisfy $\Phi(0) = 1$, then $\Phi(t\xi) \to 1$ as $t \to 0$, and with any luck the corresponding integral will converge to f in some sense. To make this precise, we need the following lemma, which complements Theorem (8.22c).

(8.30) Lemma. If $f, g \in L^2$, then $(\hat{f}\hat{g})^{\vee} = f * g$.

Proof. $\hat{f}\hat{g} \in L^1$ by Hölder's inequality, so $(\hat{f}\hat{g})^{\vee}$ makes sense. Given $x \in \mathbf{R}^n$, let $h(y) = \overline{g(x-y)}$. Then it is easily verified that $\hat{h}(\xi) = \overline{\hat{g}(\xi)}e^{-2\pi i x\cdot\xi}$, so since $\mathscr{F}$ is unitary on L^2,

$$f * g(x) = \int f\bar{h} = \int \hat{f}\bar{\hat{h}} = \int \hat{f}(\xi)\hat{g}(\xi)e^{2\pi i x\cdot\xi}\,d\xi = (\hat{f}\hat{g})^{\vee}(x). \qquad \square$$

(8.31) Theorem. Suppose $\Phi \in L^1 \cap C_0$, $\Phi(0) = 1$, and $\varphi = \check{\Phi} \in L^1$. Suppose also that $f \in L^1 + L^2$, and for $t > 0$ set

$$F^t(x) = \int \hat{f}(\xi)\Phi(t\xi)e^{2\pi i x\cdot\xi}\,d\xi.$$

(a) If $f \in L^p$ $(1 \le p < \infty)$, then $F^t \in L^p$ and $\|F^t - f\|_p \to 0$ as $t \to 0$.

(b) If f is bounded and uniformly continuous, then so is F^t, and $\|F^t - f\|_\infty \to 0$ as $t \to 0$.

(c) Suppose also that $|\varphi(x)| \le C(1 + |x|)^{-n-\varepsilon}$ for some $C, \varepsilon > 0$. Then $F^t(x) \to f(x)$ for every x in the Lebesgue set of f.

Proof. We have $f = f_1 + f_2$ where $f_1 \in L^1$, $f_2 \in L^2$. Since $\hat{f}_1 \in L^\infty$, $\hat{f}_2 \in L^2$, and $\Phi \in (L^1 \cap C_0) \subset (L^1 \cap L^2)$, the integral defining $F^t(x)$ converges absolutely for every x. Moreover, $\Phi(t\xi) = \hat{\varphi}_t(\xi)$ by Theorem (8.22b), and $\int \varphi(x)\,dx = \Phi(0) = 1$. Since $\varphi, \Phi \in L^1$ we have $f_1 * \varphi$, $\hat{f}_1\Phi \in L^1$, so by Theorem (8.22c) and the inversion formula,

$$\int \hat{f}_1(\xi)\Phi(t\xi)e^{2\pi i x\cdot\xi}\,d\xi = f_1 * \varphi_t(x).$$

Also, $\varphi \in L^2$ by the Plancherel theorem, so by Lemma (8.30),

$$\int \hat{f}_2(\xi)\Phi(t\xi)e^{2\pi i x\cdot\xi}\,d\xi = f_2 * \varphi_t(x).$$

In short, $F^t = f * \varphi_t$, so the assertions follow from Theorems 8.14 and 8.15. □

We conclude with some examples of functions Φ which satisfy the hypotheses of Theorem (8.31).

(8.32) Let $\Phi(\xi) = e^{-\pi|\xi|^2}$. Then $\varphi(x) = e^{-\pi|x|^2}$ by Proposition (8.24). φ is called the **Gaussian kernel** or **Weierstrass kernel**.

(8.33) Let $\Phi(\xi) = e^{-2\pi|\xi|}$. Then $\varphi = \check{\Phi}$ is called the **Poisson kernel**. For $n = 1$, we have

$$\begin{aligned}\varphi(x) &= \int_{-\infty}^{0} e^{2\pi(1+ix)\xi}\,d\xi + \int_{0}^{\infty} e^{2\pi(-1+ix)\xi}\,d\xi \\ &= \frac{1}{2\pi}\left[\frac{1}{1+ix} + \frac{1}{1-ix}\right] \\ &= \frac{1}{\pi(1+x^2)}.\end{aligned}$$

The formula for φ in higher dimensions is worked out in Exercise 13.

(8.34) (For $n = 1$) Let $\Phi(\xi) = \max((1 - |\xi|), 0)$. Then

$$\begin{aligned}\varphi(x) &= \int_{-1}^{0}(1+\xi)e^{2\pi ix\xi}\,d\xi + \int_{0}^{1}(1-\xi)e^{2\pi ix\xi}\,d\xi \\ &= \frac{e^{2\pi ix} + e^{-2\pi ix} - 2}{(2\pi ix)^2} \\ &= \frac{1-\cos 2\pi x}{2\pi^2 x^2} \\ &= \left(\frac{\sin \pi x}{\pi x}\right)^2.\end{aligned}$$

φ is called the **Fejér kernel**.

EXERCISES

12. Work out the analog of Theorem (8.22) for the Fourier transform on $\mathbf{T}^n$.

13. The purpose of this exercise is to prove that the inverse Fourier transform of $e^{-2\pi|\xi|}$ is

$$\varphi(x) = \frac{\Gamma((n+1)/2)}{\pi^{(n+1)/2}}(1+|x|^2)^{-(n+1)/2}.$$

(a) If $\beta \geq 0$, $e^{-\beta} = \pi^{-1}\int_{-\infty}^{\infty}(1+t^2)^{-1}e^{-i\beta t}\,dt$. [Use (8.33).]

(b) If $\beta \geq 0$, $e^{-\beta} = \int_0^\infty (\pi s)^{-1/2}e^{-s}e^{-\beta^2/4s}\,ds$. [Use (a), Proposition (8.24), and the formula $(1+t^2)^{-1} = \int_0^\infty e^{-(1+t^2)s}\,ds$.]

(c) If we set $\beta = 2\pi|\xi|$ where $\xi \in \mathbf{R}^n$, the formula in (b) expresses $e^{-2\pi|\xi|}$ as a weighted average of dilated Gaussian kernels. Use Proposition (8.24) again to derive the asserted formula for φ.

14. In this exercise, $n = 1$.

(a) If $f = \chi_{(-a,a)}$, then $\check{f}(x) = (\pi x)^{-1}\sin 2\pi ax$.

(b) Let $f_k = \chi_{(-k,k)} * \chi_{(-1,1)}$. Compute f_k explicitly and show that $\|f_k\|_\infty = 2$.

(c) $\check{f}_k(x) = (\pi x)^{-2}\sin 2\pi kx \sin 2\pi x$, and $\|\check{f}_k\|_1 \to \infty$ as $k \to \infty$. (Hint: set $y = 2\pi kx$.)

(d) $\mathscr{F}(L^1)$ is a proper subset of C_0. (Consider $g_k = \check{f}_k$ and use the open mapping theorem.)

15. If $f \in L^1(\mathbf{R}^n)$, f is continuous at 0, and $\hat{f} \geq 0$, then $\hat{f} \in L^1$. [Use Theorem (8.31c) and Fatou's lemma.]

16. Let $f(x) = e^{-2\pi x}x^{\alpha-1}$ for $x > 0$ and $f(x) = 0$ for $x \leq 0$, where $\alpha > 0$. Then $f \in L^1$ and $\hat{f}(\xi) = \Gamma(\alpha)[(2\pi)(1+i\xi)]^{-\alpha}$. [Here we are using the branch of z^α in the right half plane which is positive when z is positive. The complex substitution $y = (1+i\xi)x$ in the integral defining $\hat{f}$ is justified by Cauchy's theorem.] Deduce from this that if $a > \frac{1}{2}$ and $b > \frac{1}{2}$, then

$$\int_{-\infty}^{\infty}(1-ix)^{-a}(1+ix)^{-b}dx = \frac{2^{2-a-b}\pi\Gamma(a+b-1)}{\Gamma(a)\Gamma(b)}.$$

17. Let $f(x) = \frac{1}{2} - x$ on the interval $[0,1)$, and extend f to be periodic on $\mathbf{R}$.

(a) $\hat{f}(0) = 0$, and $\hat{f}(\kappa) = (2\pi i\kappa)^{-1}$ if $\kappa \neq 0$.

(b) $\sum_1^\infty k^{-2} = \pi^2/6$. (Use the Parseval equation.)

18. (Wirtinger's Inequality) If $f \in C^1([a,b])$ and $f(a) = f(b) = 0$, then

$$\int_a^b |f(x)|^2\,dx \leq \left(\frac{b-a}{\pi}\right)^2 \int_a^b |f'(x)|^2\,dx.$$

[By a change of variable it suffices to assume $a = 0$, $b = \frac{1}{2}$. Extend f to $[-\frac{1}{2}, \frac{1}{2}]$ by setting $f(-x) = -f(x)$, and then extend f to $\mathbf{R}$ by periodicity. Check that f, thus extended, is in $C^1(\mathbf{T}^1)$ and apply the Parseval equation.]

19. Since $\mathscr{F}$ commutes with rotations, the Fourier transform of a radial function is radial; that is, if $F \in L^1$ and $F(x) = f(|x|)$, then $\hat{F}(\xi) = g(|\xi|)$, where f and g are related as follows.

(a) Let $J(\xi) = \int_S e^{ix\cdot\xi}d\sigma(x)$ where σ is surface measure on the unit sphere S in $\mathbf{R}^n$ [Theorem (2.49)]. Then J is radial—say, $J(\xi) = j(|\xi|)$—and $g(\rho) = \int_0^\infty j(2\pi r\rho)f(r)r^{n-1}\,dr$.

(b) J satisfies $\sum_1^n \partial_k^2 J + J = 0$.

(c) j satisfies $\rho j''(\rho) + (n-1)j'(\rho) + \rho j(\rho) = 0$. [This equation is a variant of Bessel's equation. j is completely determined by the fact that it is a solution of this equation, is smooth at $\rho = 0$, and satisfies $j(0) = \sigma(S) = 2\pi^{n/2}/\Gamma(n/2)$. In fact, $j(\rho) = (2\pi)^{n/2}\rho^{(2-n)/2}J_{(n-2)/2}(\rho)$ where J_α is the Bessel function of the first kind of order α.]

(d) If $n = 3$, $j(\rho) = 4\pi(\sin\rho)/\rho$. [Set $j(\rho) = f(\rho)/\rho$ and use (c) to show that $f'' + f = 0$.]

20. In this exercise we develop the theory of Hermite functions.

(a) Define operators T, T^* on $\mathscr{S}(\mathbf{R})$ by $Tf(x) = 2^{-1/2}[xf(x) - f'(x)]$, $T^*f(x) = 2^{-1/2}[xf(x) + f'(x)]$. Then $\int (Tf)\bar{g} = \int f\overline{(T^*g)}$ and $T^*T^k - T^kT^* = kT^{k-1}$. ($T$ and T^* are called creation and annihilation operators in quantum field theory.)

(b) Let $h_0(x) = \pi^{-1/4}e^{-x^2/2}$ and, for $k \geq 1$, $h_k = (k!)^{-1/2}T^kh_0$. h_k is the kth **normalized Hermite function**. We have $Th_k = \sqrt{k+1}\,h_{k+1}$ and $T^*h_k = \sqrt{k}\,h_{k-1}$, hence $TT^*h_k = kh_k$. [The operator $S = 2TT^* + I$ is called the **Hermite operator**. We have $Sf(x) = x^2f(x) - f''(x)$ and $Sh_k = (2k+1)h_k$.]

(c) $\{h_k\}_0^\infty$ is an orthonormal set in $L^2(\mathbf{R})$. [Check directly that $\|h_0\|_2 = 1$, then observe that for $k > 0$, $\int h_k\bar{h}_m = k^{-1}\int (TT^*h_k)\bar{h}_m$ and use (a) and (b).]

(d) We have

$$T^kf(x) = (-1)^k2^{-k/2}e^{x^2/2}\left(\frac{d}{dx}\right)^k\left[e^{-x^2/2}f(x)\right]$$

(use induction on k), and in particular,

$$h_k(x) = \frac{(-1)^k}{[\pi^{1/2}2^kk!]^{1/2}}e^{x^2/2}\left(\frac{d}{dx}\right)^ke^{-x^2}.$$

(e) Let $H_k(x) = h_k(x)e^{x^2/2}$. Then H_k is a polynomial of degree precisely k, the kth **normalized Hermite polynomial**. The linear span of $H_0, \ldots, H_m$ is the set of all polynomials of degree $\leq m$.

(f) $\{h_k\}_0^\infty$ is an orthonormal basis for $L^2(\mathbf{R})$. [Suppose $f \perp h_k$ for all k, and let $g(x) = f(x)e^{-x^2/2}$. Show that $\hat{g} = 0$ by expanding $e^{-2\pi ix\xi}$ in its Maclaurin series and using (e).]

(g) Define A: $L^2 \to L^2$ by $Af(x) = (2\pi)^{1/4}f(x\sqrt{2\pi})$, and define $\tilde{f} = A^{-1}\mathscr{F}Af$ for $f \in L^2$. Then A is unitary and $\tilde{f}(\xi) = (2\pi)^{-1/2}\int f(x)e^{-ix\xi}\,dx$. Moreover, $(Tf)^\sim = -iT(\tilde{f})$ for $f \in \mathscr{S}$, and $\tilde{h}_0 = h_0$; hence $\tilde{h}_k = (-i)^kh_k$. Therefore, if $\varphi_k = Ah_k$, $\{\varphi_k\}_0^\infty$ is an orthonormal basis for L^2 consisting of eigenfunctions for $\mathscr{F}$: namely, $\hat{\varphi}_k = (-i)^k\varphi_k$.

4. FOURIER INVERSION ON THE TORUS

We begin this section by discussing a remarkable connection between the Fourier transforms on $\mathbf{R}^n$ and $\mathbf{T}^n$. Consider the following problem: given a function $f \in L^1(\mathbf{R}^n)$, how can one manufacture a periodic function (that is, a function on $\mathbf{T}^n$) from it? Two possible answers suggest themselves. One way is to "average" f over all periods, producing the series $\sum_{k\in\mathbf{Z}^n}f(x-k)$. This series, if it converges, will surely define a periodic function. The other way is to restrict $\hat{f}$ to the lattice $\mathbf{Z}^n$ and use it to form a Fourier series $\sum_{\kappa\in\mathbf{Z}^n}\hat{f}(\kappa)e^{2\pi i\kappa\cdot x}$. The content of our first theorem is that these methods both work and both give the same answer. In what follows, for purposes of integration we identify $\mathbf{T}^n$ with the cube $[-\frac{1}{2}, \frac{1}{2})^n$.

(8.35) Theorem. If $f \in L^1(\mathbf{R}^n)$, the series $\sum_{k\in\mathbf{Z}^n}f^k$ [with notation as in (8.3): $f^k(x) = f(x-k)$] converges in $L^1(\mathbf{T}^n)$ to a function Pf such that $\|Pf\|_1 \leq \|f\|_1$. Moreover, for $\kappa \in \mathbf{Z}^n$, $(Pf)\hat{}(\kappa)$ (Fourier transform on $\mathbf{T}^n$) equals $\hat{f}(\kappa)$ (Fourier transform on $\mathbf{R}^n$).

Proof. If $Q = [-\frac{1}{2}, \frac{1}{2})^n$, we have

$$\int_Q \sum_{k\in\mathbf{Z}^n}|f(x-k)|\,dx = \sum_{k\in\mathbf{Z}^n}\int_{Q+k}|f(x)|\,dx = \int_{\mathbf{R}^n}|f(x)|\,dx,$$

because $\mathbf{R}^n$ is the disjoint union of the cubes $Q + k$, $k \in \mathbf{Z}^n$. This proves the first assertion since L^1 is complete, and the second follows from Theorem (2.26):

$$\begin{aligned}(Pf)\hat{}(\kappa) &= \int_Q \sum_{k\in\mathbf{Z}^n} f(x-k)e^{-2\pi i\kappa\cdot x}\,dx \\ &= \sum_{k\in\mathbf{Z}^n}\int_{Q+k} f(x)e^{-2\pi i\kappa\cdot(x+k)}\,dx \\ &= \sum_{k\in\mathbf{Z}^n}\int_{Q+k} f(x)e^{-2\pi i\kappa\cdot x}\,dx \\ &= \int_{\mathbf{R}^n} f(x)e^{-2\pi i\kappa\cdot x}\,dx = \hat{f}(\kappa).\end{aligned}$$

□

If we now impose conditions on f to guarantee that the series in question converge absolutely, we obtain a more refined result.

(8.36) The Poisson Summation Formula. Suppose $f \in C(\mathbf{R}^n)$, $|f(x)| \leq C(1+|x|)^{-n-\varepsilon}$, and $|\hat{f}(\xi)| \leq C(1+|\xi|)^{-n-\varepsilon}$ for some $C, \varepsilon > 0$. Then

$$\sum_{k\in\mathbf{Z}^n} f(x+k) = \sum_{\kappa\in\mathbf{Z}^n}\hat{f}(\kappa)e^{2\pi i\kappa\cdot x},$$

where both sides converge absolutely and uniformly on $\mathbf{T}^n$. In particular, taking $x = 0$,

$$\sum_{k\in\mathbf{Z}^n} f(k) = \sum_{\kappa\in\mathbf{Z}^n} \hat{f}(\kappa).$$

Proof. The absolute and uniform convergence of the series follows from the fact that $\Sigma_{k\in\mathbf{Z}^n}(1+|k|)^{-n-\varepsilon} < \infty$, which can be seen by comparing the latter series to the convergent integral $\int_{\mathbf{R}^n}(1+|x|)^{-n-\varepsilon}\,dx$. Thus the function $Pf(x) = \Sigma_k f(x+k)$ is in $C(\mathbf{T}^n)$, and hence in $L^2(\mathbf{T}^n)$, so Theorem (8.35) implies that the series $\Sigma_\kappa \hat{f}(\kappa)e^{2\pi i\kappa\cdot x}$ converges in $L^2(\mathbf{T}^n)$ to Pf. Since it also converges uniformly, its sum equals Pf pointwise. □

The Poisson summation formula has a number of interesting applications. Here we are mainly interested in using it to convert the results of the previous section into methods for recovering a function on $\mathbf{T}^n$ from its Fourier transform.

(8.37) Theorem. Suppose φ satisfies the hypotheses of the Poisson summation formula and $\int\varphi(x)dx = 1$, and let $\Phi = \hat{\varphi}$. If $f \in L^1(\mathbf{T}^n)$ and $t > 0$, set

$$F^t(x) = \sum_{\kappa\in\mathbf{Z}^n} \hat{f}(\kappa)\Phi(t\kappa)e^{2\pi i\kappa\cdot x}$$

[which converges absolutely since $\Sigma_\kappa|\Phi(t\kappa)| < \infty$].

(a) If $f \in L^p(\mathbf{T}^n)$ $(1 \le p < \infty)$, then $\|F^t - f\|_p \to 0$ as $t \to 0$. If $f \in C(\mathbf{T}^n)$, $\|F^t - f\|_\infty \to 0$ as $t \to 0$.

(b) $F^t(x) \to f(x)$ for every x in the Lebesgue set of f.

Proof. If φ_t is given by (8.13), φ_t still satisfies the hypotheses of the Poisson summation formula (with C depending on t), and $\Phi(t\xi) = \hat{\varphi}_t(\xi)$. Let us denote by ψ_t the periodization of φ_t:

$$\psi_t(x) = \sum_{k\in\mathbf{Z}^n} \varphi_t(x-k) = \sum_{\kappa\in\mathbf{Z}^n} \Phi(t\kappa)e^{2\pi i\kappa\cdot x}.$$

Then $(f*\psi_t)\hat{}\,(x) = \hat{f}(\kappa)\Phi(t\kappa)$, so $F^t(x) = f*\psi_t(x)$. Since Φ is continuous and $\Phi(0) = \int\varphi = 1$, we have $\Phi(t\kappa) \to 1$ as $t \to 0$, and it follows that $f*\psi_t \to f$ uniformly [and hence also in $L^p(\mathbf{T}^n)$] if f is a trigonometric polynomial [that is, if $\hat{f}(\kappa) = 0$ for all but finitely many κ]. On the other hand, by Young's inequality and Theorem (8.33) we have

$$\|F^t\|_p \le \|f\|_p\|\psi_t\|_1 \le \|f\|_p\|\varphi_t\|_1 = \|f\|_p\|\varphi\|_1,$$

so the operators $f \to F^t$ are bounded on L^p uniformly in t, $1 \le p \le \infty$. Since

the trigonometric polynomials are uniformly dense in $C(\mathbf{T}^n)$ by the Stone–Weierstrass theorem and are thus also dense in $L^p(\mathbf{T}^n)$ $(1 \le p < \infty)$, part (a) follows from Proposition (5.17).

To prove (b), suppose x is in the Lebesgue set of f; by translating f we may assume that $x = 0$ (which simplifies the notation). With $Q = [-\frac{1}{2}, \frac{1}{2})^n$, we have

$$F^t(0) = f*\psi_t(0) = \int_Q f(x)\psi_t(-x)\,dx$$

$$= \int_Q f(x)\varphi_t(-x)\,dx + \sum_{k\neq 0}\int_Q f(x)\varphi_t(-x+k)\,dx.$$

Since

$$|\varphi_t(x)| \le Ct^{-n}(1+t^{-1}|x|)^{-n-\varepsilon} \le Ct^\varepsilon|x|^{-n-\varepsilon},$$

for $x \in Q$ and $k \neq 0$ we have

$$|\varphi_t(-x+k)| \le C2^{n+\varepsilon}t^\varepsilon|k|^{-n-\varepsilon}$$

and hence

$$\sum_{k\neq 0}\int_Q |f(x)\varphi_t(-x+k)|\,dx \le \left[C2^{n+\varepsilon}\|f\|_1\sum_{k\neq 0}|k|^{-n-\varepsilon}\right]t^\varepsilon,$$

which vanishes as $t \to 0$. On the other hand, if we define $g = f\chi_Q \in L^1(\mathbf{R}^n)$, then 0 is in the Lebesgue set of g (because 0 is in the interior of Q, and the condition that 0 be in the Lebesgue set of g depends only on the behavior of g near 0), so by Theorem (8.15),

$$\lim_{t\to 0}\int_Q f(x)\varphi_t(-x)\,dx = \lim_{t\to 0} g*\varphi_t(0) = g(0) = f(0). \qquad \square$$

If we apply this theorem to the examples (8.33) and (8.34), we obtain the classical results on summability of one-dimensional Fourier series. Some terminology: if $f \in L^1(\mathbf{T}^1)$ and $m = 0, 1, 2, \ldots$, $S_m f$ will denote the mth (symmetric) partial sum of the Fourier series of f:

$$S_m f(x) = \sum_{-m}^{m} \hat{f}(k)e^{2\pi ikx}.$$

If $\Sigma_0^\infty a_k$ is any numerical series, the series $\Sigma_0^\infty r^k a_k$ is called its rth **Abel mean**, and the average of its first $m+1$ partial sums,

$$\frac{1}{m+1}\sum_{k=0}^{m}\sum_{j=0}^{k} a_j,$$

is called its mth **Cesàro mean**. It is easily verified that if $\sum_0^\infty a_k$ converges to the sum S, then its Abel and Cesàro means converge as $r \to 1$ or $m \to \infty$ to S, but the latter limits may exist even when $\sum_0^\infty a_k$ diverges. (In applying these concepts to Fourier series, whose index of summation runs from $-\infty$ to ∞, we group together the kth and $(-k)$th terms, as with the partial sums.)

Let $\Phi(\xi) = e^{-2\pi|\xi|}$ as in (8.33). If we make the change of variable $r = e^{-2\pi t}$ in theorem (8.37), $F^t(x)$ becomes the rth Abel mean of the Fourier series of f:

$$F^t(x) = A_r f(x) = \sum_{k=0}^{\infty} r^k \hat{f}(k) e^{2\pi i k x} + \sum_{k=1}^{\infty} r^k \hat{f}(-k) e^{-2\pi i k x}.$$

The Abel means have the following complex-variable interpretation: if we set $z = re^{2\pi i x}$, then

$$A_r f(x) = \sum_0^\infty \hat{f}(k) z^k + \sum_1^\infty \hat{f}(-k) \bar{z}^k.$$

The two series on the right define a holomorphic and an antiholomorphic function on the unit disc $|z| < 1$. In particular, $A_r f(x)$ is a harmonic function on the disc, and the fact that $A_r f \to f$ as $r \to 1$ means that f is the boundary value of this function on the unit circle. (See also Exercise 22.)

Now let $\Phi(\xi) = \max(1 - |\xi|, 0)$ as in (8.34). If we take $t = (m+1)^{-1}$ and set $\sigma_m f(x) = F^{(m+1)^{-1}}(x)$ in Theorem (8.37), we have

$$\sigma_m f(x) = \sum_{-m}^{m} \hat{f}(k)\left(1 - \frac{|k|}{m+1}\right) e^{2\pi i k x}. \tag{8.38}$$

It is easily verified that $\sigma_m f$ is the mth Cesàro mean of the Fourier series of f; see Exercise 23.

The remainder of this section is devoted to a discussion of the pointwise convergence of one-dimensional Fourier series, that is, the convergence of the partial sums $S_m f(x)$. The behavior of $S_m f$ is a remarkably tricky business—much more so than that of the Cesàro or Abel means—and it is not our intention here to go into detail. Rather, we shall present a group of results which gives a reasonably complete picture for piecewise smooth functions. (See also Exercise 25.)

We begin with a simple criterion for the Fourier series of a function f on $\mathbf{T}^1$ to have an absolutely and uniformly convergent Fourier series, that is, for $\hat{f}$ to be in $l^1(\mathbf{Z})$. (Another one is given in Exercise 26.) For $0 < \alpha \leq 1$, let $\Lambda_\alpha(\mathbf{T}^1)$ be the space of functions on $\mathbf{T}^1$ which are Hölder continuous of exponent α:

$$\Lambda_\alpha(\mathbf{T}^1) = \left\{ f \in C(\mathbf{T}^1) : \sup_{x,h} \frac{|f(x+h) - f(x)|}{|h|^\alpha} < \infty \right\}.$$

[$\Lambda_\alpha(\mathbf{T}^1)$ includes, for example, all absolutely continuous functions f such that $f' \in L^\infty$: cf. Exercise 3.37.]

(8.39) Bernstein's Theorem. If $f \in \Lambda_\alpha(\mathbf{T}^1)$ where $\alpha > \frac{1}{2}$, then $\hat{f} \in l^1(\mathbf{Z})$.

Proof. Given $h \in \mathbf{R}$, let $g(x) = f(x+h) - f(x-h)$. Then

$$\hat{g}(k) = \hat{f}(k)[e^{2\pi i k h} - e^{-2\pi i k h}] = 2i\hat{f}(k)\sin 2\pi k h,$$

so by the Parseval equation,

$$\sum_{-\infty}^{\infty} |\hat{f}(k)\sin 2\pi k h|^2 = \tfrac{1}{4}\int_0^1 |f(x+h) - f(x-h)|^2\, dx \leq C_0 |h|^{2\alpha}$$

for some $C_0 > 0$. For $\nu = 0, 1, 2, \ldots$, if $2^\nu \leq |k| < 2^{\nu+1}$ and we take $h = 2^{-\nu-3}$, then

$$\frac{\pi}{4} \leq 2\pi |k| h \leq \frac{\pi}{2}, \quad \text{hence } |\sin 2\pi k h|^2 \geq \frac{1}{2}.$$

Therefore,

$$\sum_{2^\nu \leq |k| < 2^{\nu+1}} |\hat{f}(k)|^2 \leq C_1 2^{-2\alpha\nu},$$

so by the Schwarz inequality,

$$\sum_{2^\nu \leq |k| < 2^{\nu+1}} |\hat{f}(k)| \leq \left(C_1 2^{-2\alpha\nu}\right)^{1/2} (2^{\nu+1} - 2^\nu)^{1/2} \leq C_2 2^{[(1/2)-\alpha]\nu}.$$

Finally, summing over ν, we obtain

$$\sum_{-\infty}^{\infty} |\hat{f}(k)| \leq |\hat{f}(0)| + C_2 \sum_{\nu=0}^{\infty} 2^{[(1/2)-\alpha]\nu} < \infty. \qquad \square$$

We next aim to show that $S_m f \to f$ on any interval where f is smooth even when f is not globally smooth. For this purpose we need to examine the partial sum operators S_m more closely. We have

$$S_m f(x) = \sum_{-m}^{m} \int_0^1 f(y) e^{2\pi i k (x-y)}\, dy = f * D_m(x)$$

where D_m is the mth **Dirichlet kernel**:

$$D_m(x) = \sum_{-m}^{m} e^{2\pi i k x}.$$

(8.40) Lemma. $D_m(x) = \sin(2m+1)\pi x / \sin \pi x$.

Proof. By the formula for summing a finite geometric series,

$$D_m(x) = e^{-2\pi i m x} \sum_0^{2m} e^{2\pi i k x} = e^{-2\pi i m x} \frac{e^{2\pi i (2m+1)x} - 1}{e^{2\pi i x} - 1}$$

$$= \frac{e^{\pi i (2m+1)x} - e^{-\pi i (2m+1)x}}{e^{\pi i x} - e^{-\pi i x}} = \frac{\sin(2m+1)\pi x}{\sin \pi x}. \qquad \square$$

(8.41) The Localization Principle. If $f \in L^1(\mathbf{T}^1)$ and $f = 0$ on an open interval I, then $S_m f \to 0$ uniformly on compact subsets of I.

Proof. Without loss of generality we may assume that I is centered at 0, say $I = (-c, c)$. If $\delta < c$ and $|x| < \delta$, let

$$f_x^+(y) = \frac{f(x-y)e^{\pi i y}}{2i \sin \pi y}, \qquad f_x^-(y) = \frac{f(x-y)e^{-\pi i y}}{2i \sin \pi y}.$$

Since $f(x-y) = 0$ in a neighborhood of the zeros of $\sin \pi y$, we have f_x^+, $f_x^- \in L^1$, and by Lemma (8.40),

$$S_m f(x) = \int_0^1 f(x-y) D_m(y)\, dy = \int_0^1 f(x-y) \frac{e^{(2m+1)\pi i y} - e^{-(2m+1)\pi i y}}{2i \sin \pi y} dy$$

$$= (f_x^+)\hat{}\,(-m) - (f_x^-)\hat{}\,(m).$$

By the periodic version of Theorem (8.22f), therefore, $S_m f(x) \to 0$ as $m \to \infty$ for $|x| < \delta$. Moreover, if $|x_1| < \delta$ and $|x_2| < \delta$,

$$S_m f(x_1) - S_m f(x_2) = \int_0^1 \frac{\sin(2m+1)\pi y}{\sin \pi y} [f(x_1 - y) - f(x_2 - y)]\, dy.$$

But $f(x_1 - y) - f(x_2 - y) = 0$ on the complement in $[0,1]$ of $[c - \delta, 1 - (c - \delta)]$, and on the latter interval we have

$$\left| \frac{\sin(2m+1)\pi y}{\sin \pi y} \right| \le \frac{1}{\sin \pi (c - \delta)} = A$$

where A is independent of m. Hence

$$|S_m f(x_1) - S_m f(x_2)| \le A \int_0^1 |f(x_1 - y) - f(x_2 - y)|\, dy = A \| f^{x_1} - f^{x_2} \|_1,$$

which vanishes as $x_1 - x_2 \to 0$ by (the periodic analog of) Proposition (8.5). Therefore, given $\varepsilon > 0$, we can choose η small enough so that if $|x_1| < \delta$, $|x_2| < \delta$, and $|x_1 - x_2| < \eta$, we have $|S_m f(x_1) - S_m f(x_2)| < \varepsilon/2$. Choose $x_1, \ldots, x_k \in [-\delta, \delta]$ so that the intervals $|x - x_j| < \eta$ cover $[-\delta, \delta]$. Since $S_m f(x_j) \to 0$ as we showed above, we can choose M large enough so that $|S_m f(x_j)| < \varepsilon/2$ for $m > M$ and $1 \le j \le k$. If $|x| < \delta$, then, we have $|x - x_j| < \eta$ for some j, so

$$|S_m f(x)| \le |S_m f(x) - S_m f(x_j)| + |S_m f(x_j)| < \varepsilon$$

for $m > M$, and we are done. $\square$

(8.42) Corollary. Suppose $f \in L^1(\mathbf{T}^1)$.

(a) If $g \in L^1(\mathbf{T}^1)$ and $f = g$ on an open interval I, then $S_m f - S_m g \to 0$ uniformly on compact subsets of I.

(b) If f agrees with a function $g \in \Lambda_\alpha(\mathbf{T}^1)$ on an open interval I, where $\alpha > \frac{1}{2}$, then $S_m f \to f$ uniformly on compact subsets of I.

Proof. For (a), apply the localization principle to $f - g$. (b) then follows since $S_m g \to g$ uniformly by Bernstein's theorem. $\square$

Finally, we discuss the behavior of $S_m f$ near a jump discontinuity of f. Let us first consider a simple example: let $\varphi(x) = \frac{1}{2} - x$ for $0 \le x < 1$, and extend φ to be periodic on $\mathbf{R}$. Then φ is C^∞ except for jump discontinuities at the integers, where $\varphi(j^+) - \varphi(j^-) = 1$. It is easy to check that $\hat{\varphi}(0) = 0$ and $\hat{\varphi}(k) = (2\pi i k)^{-1}$ for $k \neq 0$ (Exercise 17), so that

$$S_m \varphi(x) = \sum_{0 < |k| \le m} \frac{e^{2\pi i k x}}{2\pi i k} = \sum_1^m \frac{\sin 2\pi k x}{\pi k}.$$

From Corollary (8.42) it follows that $S_m \varphi \to \varphi$ uniformly on any compact set not containing an integer; also, it is clear that $S_m \varphi(j) = 0$ for any $m \in \mathbf{N}$, $j \in \mathbf{Z}$. But near the discontinuities a peculiar thing happens: $S_m \varphi$ overshoots and undershoots φ, as shown in the accompanying figure.

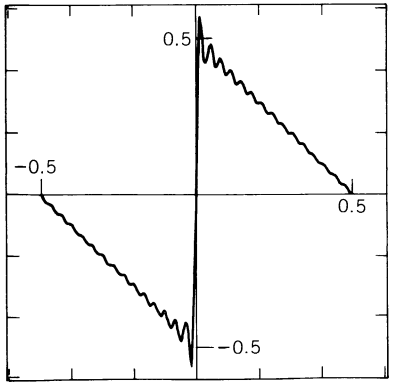

Figure 1. $y = \sum_1^{30} \frac{\sin 2\pi k x}{\pi k}$, $-0.5 < x < 0.5$.

As $m \to \infty$ the overshoots tend to zero in width but *not* in height. In fact, the amount of the first overshoot is about 18% in the limit (a figure which is given incorrectly in many places in the literature). This is known as the **Gibbs phenomenon**; for the precise statement and proof, see Exercise 28.

Now suppose f is any periodic function on $\mathbf{R}$ having a jump discontinuity at $x = a$ [i.e., $f(a^+)$ and $f(a^+)$ exist and are unequal]. Then the function

$$g(x) = f(x) - [f(a^+) - f(a^-)]\varphi(x - a) \quad (x \neq a)$$

is continuous at every point where f is, and also at $x = a$ provided we define $g(a) = \frac{1}{2}[f(a^+) + f(a^-)]$. With this in mind, let us define a function f on $\mathbf{T}^1$ to be **piecewise regular** if f is continuous on $\mathbf{T}^1$ except for a finite number of jump discontinuities at $a_1, \dots, a_k$, and the function g given by

$$g(x) = f(x) - \sum_1^k \left[f(a_j^+) - f(a_j^-)\right]\varphi(x - a_j) \quad (x \neq a_1, \dots, a_k),$$

$$g(a_i) = \tfrac{1}{2}\left[f(a_i^+) + f(a_i^-)\right] - \sum_{j \neq i} \left[f(a_j^+) - f(a_j^-)\right]\varphi(a_i - a_j)$$

(which is continuous everywhere) lies in $\Lambda_\alpha(\mathbf{T}^1)$ for some $\alpha > \frac{1}{2}$. Then f exhibits the Gibbs phenomenon at each a_j since $\varphi(x - a_j)$ does so and $S_m g \to g$ uniformly. Combining this with the preceding results, we have proved:

(8.43) Theorem. Suppose f is piecewise regular on $\mathbf{T}^1$, with discontinuities at $a_1, \dots, a_k$. Then $S_m f \to f$ uniformly on any compact set disjoint from $\{a_1, \dots, a_k\}$; $S_m f(a_j) \to \frac{1}{2}[f(a_j^+) + f(a_j^-)]$ for each j, and $S_m f$ exhibits the Gibbs phenomenon at each a_j.

EXERCISES

21. Some applications of the Poisson summation formula [cf. (8.32) and (8.33)]:

(a) (Jacobi's identity) $\sum_{-\infty}^{\infty} e^{-\pi k^2/a} = \sqrt{a} \sum_{-\infty}^{\infty} e^{-\pi a k^2}$ $(a > 0)$.

(b) $\sum_{-\infty}^{\infty} (k^2 + a^2)^{-1} = \dfrac{\pi}{a} \cdot \dfrac{1 + e^{-2\pi a}}{1 - e^{-2\pi a}}$ $(a > 0)$.

(c) Subtract a^{-2} from both sides of (b) and let $a \to 0$ to obtain $\sum_1^\infty k^{-2} = \pi^2/6$.

22. Let f be a function on $\mathbf{T}^1$ and $A_r f$ the rth Abel mean of the Fourier series of f.

(a) $A_r f = f * P_r$ where $P_r(x) = \sum_{-\infty}^{\infty} r^{|k|} e^{2\pi i k x}$ is the **Poisson kernel** for $\mathbf{T}^1$.

(b) $P_r(x) = (1 - r^2)/(1 + r^2 - 2r\cos 2\pi x)$.

23. If f is a function on $\mathbf{T}^1$, let $\sigma_m f$ be given by (8.38).

(a) $\sigma_m f = (m + 1)^{-1} \sum_0^m S_k f$.

(b) If D_k is the kth Dirichlet kernel, we have $\sigma_m f = f * F_m$ where $F_m = (m + 1)^{-1} \sum_0^m D_k$. F_m is the mth **Fejér kernel** on $\mathbf{T}^1$.

(c) $F_m(x) = \sin^2(m + 1)\pi x/(m + 1)\sin^2 \pi x$. [Use Lemma (8.40) and the fact that $\sin(2k + 1)\pi x = \operatorname{Im} e^{(2k+1)\pi x}$.]

24. (a) If D_m is the mth Dirichlet kernel, $\|D_m\|_1 \to \infty$ as $m \to \infty$. (Use Exercise 2.61a.)

(b) The Fourier transform is not surjective from $L^1(\mathbf{T}^1)$ to $C_0(\mathbf{Z})$. (Cf. Exercise 14.)

25. (a) Let $\varphi_m(f) = S_m f(0)$. Then $\varphi_m \in C(\mathbf{T}^1)^*$ and $\|\varphi_m\| = \|D_m\|_1$.

(b) There is a dense G_δ set $E \subset C(\mathbf{T}^1)$ such that the sequence $\{S_m f(0)\}$ diverges for $f \in E$. (Use Exercise 24a.)

(c) Obviously the same result holds if $S_m f(0)$ is replaced by $S_m f(a)$ for any a. Use Exercise 5.44 to obtain a stronger result.

26. For $1 \leq p \leq \infty$, let $L_1^p(\mathbf{T}^1)$ be the set of all absolutely continuous functions on $\mathbf{T}^1$ such that $f' \in L^p(\mathbf{T}^1)$. If $f \in L_1^p(\mathbf{T}^1)$ where $p > 1$, then $\hat{f} \in l^1(\mathbf{Z})$. [It suffices to assume $p \leq 2$; use the equation $\hat{f}(k) = (2\pi i k)^{-1}(f')\hat{}(k)$ together with the inequalities of Hölder and Hausdorff–Young.]

27. (A comparison of Bernstein's theorem with Exercise 26.)

(a) If $p^{-1} + q^{-1} = 1$, then $L_1^p(\mathbf{T}^1) \subset \Lambda_{1/q}(\mathbf{T}^1)$. If $\alpha > q^{-1}$, then $L_1^p(\mathbf{T}^1) \not\subset \Lambda_\alpha(\mathbf{T}^1)$. [Hint: $f(b) - f(a) = \int_a^b f'(t)\,dt$.]

(b) If $\alpha < 1$, $\Lambda_\alpha(\mathbf{T}^1)$ contains functions which are not of bounded variation, and hence not in $L_1^p(\mathbf{T}^1)$. (But cf. Exercise 3.37.)

28. Let $\varphi(x) = \frac{1}{2} - x$ on $[0, 1)$, $\varphi(x + k) = \varphi(x)$ for $k \in \mathbf{Z}$, and $\Delta_m = S_m\varphi - \varphi$.

(a) $(d/dx)\Delta_m(x) = D_m(x) = \sin(2m + 1)\pi x/\sin \pi x$ for $0 < x < 1$.

(b) The first maximum of Δ_m to the right of 0 occurs at $x = (2m + 1)^{-1}$, and

$$\lim_{m \to \infty} \Delta_m\left((2m + 1)^{-1}\right) = \frac{1}{\pi}\int_0^\pi \frac{\sin t}{t}\,dt - \frac{1}{2} \cong 0.0895$$

$$= (0.179)\varphi(0^+).$$

[Hint: $\Delta_m(x) = \int_0^x \Delta_m'(t)\,dt - \frac{1}{2}$.]

(c) More generally, the successive maxima and minima of Δ_m on $(0, 1)$ occur at $x = j/(2m + 1)$ $(j = 1, \dots, 2m)$ and

$$\lim_{m \to \infty} \Delta_m\left(\frac{j}{2m + 1}\right) = \frac{1}{\pi}\int_0^{j\pi} \frac{\sin t}{t}\,dt - \frac{1}{2}.$$

These numbers are positive for j odd and negative for j even, and they tend to zero as $j \to \infty$. (Cf. Exercise 2.61b.)

29. A C^∞ function f on $\mathbf{R}$ is **real analytic** if for every $x \in \mathbf{R}$, f is the sum of its Taylor series centered at x in some neighborhood of x. If f is periodic and we regard f as a function on $S^1 = \{z \in \mathbf{C}: |z| = 1\}$, this means that f is the restriction to S^1 of a holomorphic function on some neighborhood of S^1. Show that $f \in C^\infty(\mathbf{T}^1)$ is real analytic iff $|\hat{f}(k)| \leq Ce^{-\epsilon k}$ for some $C, \epsilon > 0$. (See the discussion of Abel means in the text, and note that $\bar{z} = z^{-1}$ when $|z| = 1$.)

5. DISTRIBUTIONS

One of the most important conceptual advances in modern analysis is the theory of distributions or "generalized functions," which allow one to deal in a rigorous and systematic way with objects which, roughly speaking, resemble functions but are more singular than functions. We begin by discussing the so-called tempered distributions.

We recall from Section 1 that the Schwartz space $\mathcal{S}$ is a Fréchet space whose topology is defined by the norms $\| \ \|_{(N,\alpha)}$. A **tempered distribution** is a continuous linear functional on $\mathcal{S}$, and the set of all tempered distributions is denoted by $\mathcal{S}'$. $\mathcal{S}'$ is thus the dual of $\mathcal{S}$; we depart from our usual notation, which would be $\mathcal{S}^*$, in deference to common usage. If $F \in \mathcal{S}'$ and $\varphi \in \mathcal{S}$, we shall denote the value of F at φ by $\langle F, \varphi \rangle$. (Observe that the pairing $\langle \ , \ \rangle$ between $\mathcal{S}$ and $\mathcal{S}'$ is linear in both variables; this conflicts with our earlier notation for scalar products but will cause no serious confusion.) We endow $\mathcal{S}'$ with the weak * topology, that is, the topology generated by the functionals $F \to \langle F, \varphi \rangle$, $\varphi \in \mathcal{S}$.

Let us see some examples of tempered distributions.

(i) A measurable function f on $\mathbf{R}^n$ will be called **tempered** if $(1 + |x|)^{-N} f \in L^1$ for some $N \geq 0$. For example, every $f \in L^p$ $(1 \leq p \leq \infty)$ is tempered: this follows from Hölder's inequality, for $(1 + |x|)^{-N} \in L^q$ provided $N > n/q$. Every tempered function f defines a tempered distribution F by the formula

$$\langle F, \varphi \rangle = \int f\varphi$$

[clearly $|\langle F, \varphi \rangle| \leq C\|\varphi\|_{(N,0)}$ if $(1 + |x|)^{-N} f \in L^1$], and two functions define the same distribution iff they are equal a.e. We shall identify f with F, and in this way we regard the space of tempered functions as a subset of $\mathcal{S}'$. This justifies the idea that tempered distributions are generalized functions.

(ii) A positive, signed, or complex Borel measure μ on $\mathbf{R}^n$ is called **tempered** if $\int (1 + |x|)^{-N} d|\mu| < \infty$ for some $N \geq 0$. Every tempered measure μ defines a tempered distribution F by $\langle F, \varphi \rangle = \int \varphi d\mu$. This distribution F falls into category (i) above precisely when $\mu \ll m$, in which case it is the distribution defined by the function $d\mu/dm$.

(iii) For any $x \in \mathbf{R}^n$ and any multi-index α, the map F defined by $\langle F, \varphi \rangle = \partial^\alpha \varphi(x)$ is a tempered distribution. Such distributions do not arise from functions; they arise from measures precisely when $\alpha = 0$, in which case F is the point mass at x.

There is a general procedure for extending various operations from $\mathcal{S}$ to $\mathcal{S}'$. Suppose T, T' are continuous linear operators on $\mathcal{S}$ such that $\int (T\varphi)\psi = \int \varphi(T'\psi)$ for all $\varphi, \psi \in \mathcal{S}$. Then if we regard $\mathcal{S}$ as a subset of $\mathcal{S}'$ according to (i) above, we have $\langle T\varphi, \psi \rangle = \langle \varphi, T'\psi \rangle$. We therefore extend T to $\mathcal{S}'$ by defining

$$\langle TF, \varphi \rangle = \langle F, T'\varphi \rangle \quad (F \in \mathcal{S}', \varphi \in \mathcal{S}).$$

This extension is clearly a continuous operator on $\mathcal{S}'$. The most important examples are the following ones. In all of them, the original T is defined not only on $\mathcal{S}$ but on a larger class of functions such as L^p; it is easily checked that the extension of T to $\mathcal{S}'$ agrees with the original T also on this larger class.

(1) (Multiplication by smooth functions) A C^∞ function f on $\mathbf{R}^n$ is called **slowly increasing** if for every multi-index α there is an integer $N \geq 0$ such that $(1 + |x|)^{-N} \partial^\alpha f$ is bounded. We observe in passing that if f is slowly increasing then $\partial^\alpha f$ is tempered for all α, for

$$(1 + |x|)^{-N} \partial^\alpha f \in L^\infty \text{ implies } (1 + |x|)^{-N-n-1} \partial^\alpha f \in L^1.$$

Every polynomial on $\mathbf{R}^n$ is slowly increasing, as are the functions $(1 + |x|^2)^{s/2}$, $s \in \mathbf{R}$. If f is slowly increasing, the map T: $\varphi \to f\varphi$ is easily seen to be continuous on $\mathcal{S}$, and $T = T'$. If $F \in \mathcal{S}'$, we therefore define $fF \in \mathcal{S}'$ by $\langle fF, \varphi \rangle = \langle F, f\varphi \rangle$.

(2) (Translation) If $y \in \mathbf{R}^n$, let $T\varphi = \varphi^y$, that is, $T\varphi(x) = \varphi(x - y)$. The estimate (8.12) implies that T is continuous on $\mathcal{S}$, and $T'\varphi = \varphi^{-y}$, so we define the translate F^y of a tempered distribution F by $\langle F^y, \varphi \rangle = \langle F, \varphi^{-y} \rangle$.

(3) (Composition with linear maps) Let S be an invertible linear transformation of $\mathbf{R}^n$, and define $T\varphi = \varphi \circ S$. Then $T'\varphi = |\det S|^{-1} \varphi \circ S^{-1}$ by Theorem (2.43), so for $F \in \mathcal{S}'$ we define $F \circ S$ by $\langle F \circ S, \varphi \rangle = |\det S|^{-1} \langle F, \varphi \circ S^{-1} \rangle$. A particular case which we shall use below is $Sx = -x$, for which we employ the notation

$$\tilde{\varphi}(x) = \varphi(-x), \qquad \langle \tilde{F}, \varphi \rangle = \langle f, \tilde{\varphi} \rangle.$$

(4) (Differentiation) Let $T\varphi = \partial^\alpha\varphi$. Then, by integration by parts, $T'\varphi = (-1)^{|\alpha|}\partial^\alpha\varphi$, so for $F \in \mathscr{S}'$ we define $\partial^\alpha F$ by $\langle \partial^\alpha F, \varphi\rangle = (-1)^{|\alpha|}\langle F, \partial^\alpha\varphi\rangle$.

(5) (Convolution, first method) If $\psi \in \mathscr{S}$, define $T\varphi = \varphi * \psi$. T is continuous on $\mathscr{S}$ by Proposition (8.11) and its proof. Moreover,

$$\langle T\varphi, \theta\rangle = \iint \varphi(y)\psi(x-y)\theta(x)\,dy\,dx = \langle \varphi, \theta * \tilde{\psi}\rangle,$$

so we define the convolution $F * \psi$ for $F \in \mathscr{S}'$, $\psi \in \mathscr{S}$ as a tempered distribution by $\langle F * \psi, \varphi\rangle = \langle F, \varphi * \tilde{\psi}\rangle$.

(6) (Convolution, second method) If $\varphi, \psi \in \mathscr{S}$, we have

$$\varphi * \psi(x) = \int \varphi(y)\psi(x-y)dy = \langle \varphi, (\tilde{\psi})^x\rangle.$$

If $F \in \mathscr{S}'$ and $\psi \in \mathscr{S}$, then, we define $F * \psi$ as a function by $F * \psi(x) = \langle F, (\tilde{\psi})^x\rangle$. This construction yields the same result as (5): see Proposition (8.44) below.

(7) (Fourier transform) Let $T\varphi = \hat{\varphi}$. Then $T = T'$ [Lemma (8.25)], so for $F \in \mathscr{S}'$ we define $\hat{F} \in \mathscr{S}'$ by $\langle \hat{F}, \varphi\rangle = \langle F, \hat{\varphi}\rangle$. It is easily verified that the basic properties of the Fourier transform continue to hold for this extension if we employ the definitions (1)–(5) above. More precisely, parts (a)–(c) of Theorem (8.22) hold verbatim, $((-2\pi ix)^\alpha F)\hat{} = \partial^\alpha\hat{F}$, $(\partial^\alpha F)\hat{} = (2\pi i\xi)^\alpha\hat{F}$, and $(\hat{F})\check{} = F$ where $\check{G}$ is defined to be $(\hat{G})\tilde{}$.

Having completed our list, we now show that (5) and (6) are equivalent.

(8.44) Proposition. If $f \in \mathscr{S}'$ and $\psi \in \mathscr{S}$, let $F * \psi(x) = \langle F, (\tilde{\psi})^x\rangle$. Then $F * \psi$ is a slowly increasing C^∞ function, and for any $\psi \in \mathscr{S}$, $\langle F * \psi, \varphi\rangle = \langle F, \varphi * \tilde{\psi}\rangle$.

Proof. Let $v_1, \ldots, v_n$ be the standard basis for $\mathbf{R}^n$. One easily sees that

$$t^{-1}\left[(\tilde{\psi})^{x+tv_j} - (\tilde{\psi})^x\right] \to \left((\partial_j\psi)\tilde{}\right)^x \text{ (convergence in } \mathscr{S}) \text{ as } t \to 0.$$

It follows that $\partial_j(F * \psi)$ exists and equals $F * \partial_j\psi$, so by induction, $F * \psi$ is C^∞. Moreover, by Proposition (5.15) there exist N, M such that

$$|\langle F, \varphi\rangle| \le \sum_{m=1}^{M} \sum_{|\alpha|\le N} C\|\varphi\|_{(m,\alpha)} \quad (\varphi \in \mathscr{S}),$$

and hence by (8.12),

$$|F * \psi(x)| \le C \sum_{m=1}^{M} \sum_{|\alpha|\le N} \sup_y (1+|y|)^m |\partial^\alpha\psi(x-y)|$$

$$\le C(1+|x|)^M \sum_m \sum_\alpha \sup_y (1+|x-y|)^m |\partial^\alpha\psi(x-y)|$$

$$\le C(1+|x|)^M \sum_m \sum_\alpha \|\psi\|_{(m,\alpha)}.$$

The same reasoning applies with ψ replaced by $\partial^\beta\psi$, so $F * \psi$ is slowly increasing. Finally, if $\varphi \in \mathscr{S}$, we have

$$\varphi * \tilde{\psi}(y) = \int \psi(x-y)\varphi(x)\,dx = \int (\tilde{\psi})^x(y)\varphi(x)\,dx,$$

that is, $\varphi * \tilde{\psi} = \int \tilde{\psi}^x\varphi(x)\,dx$. Since φ is rapidly decreasing at ∞, the latter integral can be approximated by finite Riemann sums $\Sigma(\tilde{\psi})^{x_j}\varphi(x_j)\Delta x_j$ in the topology of $\mathscr{S}$. (We leave it to the reader to make this precise.) Therefore,

$$\langle F, \varphi * \tilde{\psi}\rangle = \lim\langle F, \textstyle\sum (\tilde{\psi})^{x_j}\varphi(x_j)\,\Delta x_j\rangle = \lim \textstyle\sum \langle F, (\tilde{\psi})^{x_j}\rangle\varphi(x_j)\,\Delta x_j$$

$$= \lim \textstyle\sum F * \psi(x_j)\varphi(x_j)\,\Delta x_j = \int F * \psi(x)\varphi(x)\,dx = \langle F * \psi, \varphi\rangle.$$

□

We now prove some approximation theorems for $\mathscr{S}$ and $\mathscr{S}'$.

(8.45) Proposition. Suppose $\psi \in \mathscr{S}$, $\psi(0) = 1$, and let $\psi^\varepsilon(x) = \psi(\varepsilon x)$. Then for any $\varphi \in \mathscr{S}$, $\psi^\varepsilon\varphi \to \varphi$ in the topology of $\mathscr{S}$ as $\varepsilon \to 0$.

Proof. Given $\delta > 0$ and $N \in \mathbf{N}$, choose a compact set K such that $(1+|x|)^N|\varphi(x)| < \delta$ for $x \notin K$. Then

$$\left\|(1+|x|)^N(\psi^\varepsilon\varphi - \varphi)\right\|_u \le \delta\|\psi\|_u + \sup_K (1+|x|)^N|\varphi(x)||\psi(\varepsilon x) - 1|,$$

and the second term on the right tends to zero as $\varepsilon \to 0$. Hence $\|\psi_\varepsilon\varphi - \varphi\|_{(N,0)} \to 0$ for every N. For the norms involving derivatives, we observe that

$$|\partial^\beta\psi^\varepsilon(x)| = \varepsilon^{|\beta|}|(\partial^\beta\psi)^\varepsilon(x)| \le C_\beta\varepsilon^{|\beta|},$$

so by the product rule,

$$(1+|x|)^N\partial^\alpha(\psi^\varepsilon\varphi - \varphi) = (1+|x|)^N(\psi^\varepsilon\partial^\alpha\varphi - \partial^\alpha\varphi) + E_\varepsilon$$

where E_ε tends to zero uniformly with ε. The preceding argument then shows that $\|\psi^\varepsilon\varphi - \varphi\|_{(N,\alpha)} \to 0$. □

(8.46) Corollary. Suppose $\psi \in \mathscr{S}$ and $\int\psi(x)\,dx = 1$. If $\varphi \in \mathscr{S}$, then $\varphi * \psi_t \to \varphi$ in the topology of $\mathscr{S}$ as $t \to 0$ [where ψ_t is as in (8.13)]. If $F \in \mathscr{S}'$, then $F * \psi_t \to F$ in the topology of $\mathscr{S}'$ as $t \to 0$.

Proof. The first assertion is true in view of Corollary (8.28) and the fact that $(\varphi * \psi_t)\hat{} = \hat{\varphi}\hat{\psi}^t$. The second one follows since $\tilde{\psi}$ satisfies the same hypotheses as ψ, so for any $\varphi \in \mathscr{S}$,

$$\langle F * \psi_t, \varphi\rangle = \langle F, \varphi * \tilde{\psi}_t\rangle \to \langle F, \varphi\rangle. \qquad \square$$

(8.47) Corollary. C_c^∞ is dense in $\mathscr{S}$ (in the topology of $\mathscr{S}$) and in $\mathscr{S}'$ (in the topology of $\mathscr{S}'$).

Proof. For the first assertion, take $\psi \in C_c^\infty$ in Proposition (8.45). For the second, given $F \in \mathscr{S}'$, choose $\psi, \theta \in \mathscr{S}$ with $\int\psi = 1$ and $\theta(0) = 1$. Then $F * \psi_t \to F$ in $\mathscr{S}'$ by Corollary (8.46), and by dualizing Proposition (8.45) one sees that $(F * \psi_t)\theta^\varepsilon \to F * \psi_t$ as $\varepsilon \to 0$. But $(F * \psi_t)\theta^\varepsilon \in C_c^\infty$ by Proposition (8.44). □

Frequently it is useful to consider a larger class of generalized functions, which bears much the same relation to $\mathscr{S}'$ as L^1_{loc} does to L^1. Namely, we define a **distribution** on $\mathbf{R}^n$ to be a linear functional F on C_c^∞ such that for every compact $K \subset \mathbf{R}^n$, the restriction of F to $C_c^\infty(K)$ is continuous with respect to the topology of $\mathscr{S}$ on $C_c^\infty(K)$. [We remark that, for fixed K, this topology is defined by the norms $\|\ \|_{(0,\alpha)}$, that is, $\varphi_k \to \varphi$ in $C_c^\infty(K)$ iff $\partial^\alpha\varphi_k \to \partial^\alpha\varphi$ uniformly for all α. The reason is that if $C = \sup_{x\in K}(1 + |x|)$, we have $\|\varphi\|_{(N,\alpha)} \le C^N\|\varphi\|_{(0,\alpha)}$ for $\varphi \in C_c^\infty(K)$.] The space of all distributions is denoted by $\mathscr{D}'$ ($\mathscr{D}$ being Schwartz's notation for C_c^∞), and we endow $\mathscr{D}'$ with the weak topology induced by C_c^∞. The following characterizations of distributions are sometimes useful.

(8.48) Proposition. If F is a linear functional on C_c^∞, the following are equivalent:

(a) F is a distribution.

(b) For every compact $K \subset \mathbf{R}^n$ there exists $G_K \in \mathscr{S}'$ which agrees with F on $C_c^\infty(K)$.

(c) For every $\psi \in C_c^\infty$, the functional $\varphi \to \langle F, \psi\varphi\rangle$ ($\varphi \in \mathscr{S}$) is a tempered distribution.

Proof. If F is a distribution and $\psi \in C_c^\infty$, the map $\varphi \to \psi\varphi$ is continuous from $\mathscr{S}$ to $C_c^\infty(\mathrm{supp}(\psi))$ (by the product rule), and if F is continuous on the latter space, the map $\varphi \to \langle F, \psi\varphi\rangle$ is continuous on $\mathscr{S}$. Thus (a) implies (c). Next, given a compact $K \subset \mathbf{R}^n$, choose $\psi \in C_c^\infty$ with $\psi = 1$ on K (by the C^∞ Urysohn lemma), and set $\langle G_K, \varphi\rangle = \langle F, \psi\varphi\rangle$. Since $\psi\varphi = \varphi$ when $\mathrm{supp}(\varphi) \subset K$, it follows that (c) implies (b). Finally, that (b) implies (a) is obvious from the definition of distribution. □

We observe that if $F \in \mathscr{S}'$, then $F|C_c^\infty \in \mathscr{D}'$. Moreover, by Corollary (8.47) the restriction map $F \to F|C_c^\infty$ is injective, so we can (and do) identify $\mathscr{S}'$ with a subset of $\mathscr{D}'$. From the preceding discussion it is clear that arbitrary distributions are locally like tempered distributions, but without the global control on their behavior near infinity. For example, any locally integrable function f, or any Radon measure μ, defines a distribution.

The philosophy expounded above for defining operations on $\mathscr{S}'$ works also on $\mathscr{D}'$. That is, given a pair of operators T, T' on C_c^∞ which satisfy $\int(T\varphi)\psi = \int\varphi(T'\psi)$ and whose restrictions to $C_c^\infty(K)$ are continuous in the topology of $\mathscr{S}$ for each compact K, T can be extended to $\mathscr{D}'$ by setting $\langle TF, \varphi\rangle = \langle F, T'\varphi\rangle$. Thus distributions can be translated, differentiated, composed with linear maps, and multiplied by arbitrary C^∞ functions. Also, if $F \in \mathscr{D}'$ and $\psi \in C_c^\infty$, $F * \psi$ is a C^∞ function defined by $F * \psi(x) = \langle F, (\tilde{\psi})^x\rangle$ which, as a distribution, satisfies $\langle F * \psi, \varphi\rangle = \langle F, \varphi * \tilde{\psi}\rangle$. The one basic operation which does not extend to $\mathscr{D}'$ is the Fourier transform; this is one of the main reasons for singling out tempered distributions for special attention.

If $F \in \mathscr{D}'$ and U is open in $\mathbf{R}^n$, we say that F **vanishes** on U, or $F = 0$ on U, if $\langle F, \varphi\rangle = 0$ for all $\varphi \in C_c^\infty(U)$. Likewise, if $F, G \in \mathscr{D}'$, we say that $F = G$ on U if $F - G = 0$ on U.

(8.49) Proposition. Let $\{U_\alpha\}$ be a collection of open sets in $\mathbf{R}^n$ and let $U = \bigcup_\alpha U_\alpha$. If $F \in \mathscr{D}'$ and F vanishes on each U_α, then F vanishes on U.

Proof. If $\varphi \in C_c^\infty(U)$, there exist $\alpha_1, \dots, \alpha_m$ such that $\mathrm{supp}(\varphi) \subset \bigcup_1^m U_{\alpha_j}$. Pick $\psi_1, \dots, \psi_m \in C_c^\infty$ such that $\mathrm{supp}(\psi_j) \subset U_{\alpha_j}$ and $\sum_1^m\psi_j = 1$ on $\mathrm{supp}(\varphi)$. [That this can be done is the C^∞ analog of Proposition (4.41), proved in the same way as that result by using the C^∞ Urysohn lemma.] Then $\langle F, \varphi\rangle = \sum_1^m\langle F, \psi_j\varphi\rangle = 0$. □

According to Proposition (8.49), if $F \in \mathscr{D}'$ there is a largest open set on which F vanishes; its complement is called the **support** of F. The set of all compactly supported distributions is commonly denoted by $\mathscr{E}'$. Clearly $\mathscr{E}' \subset \mathscr{S}'$: if $F \in \mathscr{E}'$, and $\psi \in C_c^\infty$ satisfies $\psi = 1$ on $\mathrm{supp}(F)$, then $F = \psi F \in \mathscr{S}'$ by Proposition (8.48). In fact, $\mathscr{E}'$ is a dual space in its own right. Let us endow the space $\mathcal{C}^\infty$ of all smooth functions on $\mathbf{R}^n$ with the Fréchet space topology defined by the seminorms

$$\|f\|_{[m,\alpha]} = \sup_{|x|\le m}|\partial^\alpha f(x)| \quad (m \in \mathbf{N},\ \alpha \text{ a multi-index}).$$

Thus $f_k \to f$ in C^∞ iff $\partial^\alpha f_k \to \partial^\alpha f$ uniformly on compact sets for all α.

(8.50) Theorem. $\mathscr{E}'$ is the dual of C^∞, in the sense that every $F \in \mathscr{E}'$ extends uniquely to a continuous linear functional on C^∞, and every such functional arises in this way.

Proof. If $F \in \mathscr{E}'$, choose $\psi \in C_c^\infty$ with $\psi = 1$ on $\operatorname{supp}(F)$, and define the linear functional F_0 on C^∞ by $\langle F_0, \varphi \rangle = \langle F, \psi\varphi \rangle$. Clearly $F_0|C_c^\infty = F$. Let $m \in \mathbf{N}$ be such that $\operatorname{supp}(\psi) \subset \{x: |x| \le m\}$. Then since $F \in \mathscr{S}'$, there exist C, M, N such that

$$\begin{aligned} |\langle F_0, \varphi \rangle| &\le C \sum_{|\alpha| \le N} \sup_x (1 + |x|)^M |\partial^\alpha(\psi\varphi)(x)| \\ &\le (1 + m)^M \sum_{|\alpha| \le N} C_{\alpha,\psi} \sup_{|x| \le m} |\partial^\alpha \varphi(x)| \\ &\le C' \sum_{|\alpha| \le N} \|\varphi\|_{[m,\alpha]}, \end{aligned}$$

so that F_0 is continuous on C^∞. This extension of F is unique because C_c^∞ is dense in C^∞. [If $\psi \in C_c^\infty$ and $\psi(x) = 1$ for $|x| \le 1$, set $\psi^\varepsilon(x) = \psi(\varepsilon x)$: then clearly $\|\psi^\varepsilon\varphi - \varphi\|_{[m,\alpha]} = 0$ for any $\varphi \in C^\infty$ provided $\varepsilon < m^{-1}$.] Finally, if G is a continuous linear functional on C^∞, by Proposition (5.15) there exist C, m, N such that $|\langle G, \varphi \rangle| \le C\Sigma_{|\alpha| \le N} \|\varphi\|_{[m,\alpha]}$. Since $\|\varphi\|_{[m,\alpha]} \le \|\varphi\|_{(0,\alpha)}$, the restriction of G to $\mathscr{S}$ is in $\mathscr{S}'$, and since $\|\varphi\|_{[m,\alpha]} = 0$ if $\operatorname{supp}(\varphi) \subset \{x: |x| > m\}$, we have $\operatorname{supp}(G|\mathscr{S}) \subset \{x: |x| \le m\}$, so $G|\mathscr{S} \in \mathscr{E}'$. □

If $F \in \mathscr{E}'$, there are two ways to define the Fourier transform of F: as a distribution, by $\langle \hat{F}, \varphi \rangle = \langle F, \hat{\varphi} \rangle$, and as a function, by $\hat{F}(\xi) = \langle F, E_\xi \rangle$ where $E_\xi = e^{-2\pi i x \cdot \xi}$. The argument in the proof of Proposition (8.44) also shows that these two definitions coincide. To summarize it briefly: if $G(\xi) = \langle F, E_\xi \rangle$, then for any $\varphi \in \mathscr{S}$,

$$\begin{aligned} \langle G, \varphi \rangle &= \int \langle F, E_\xi \rangle \varphi(\xi)\, d\xi = \lim \sum \langle F, E_{\xi_j} \rangle \varphi(\xi_j) \Delta\xi_j \\ &= \lim \Big\langle F, \sum E_{\xi_j} \varphi(\xi_j) \Delta\xi_j \Big\rangle = \Big\langle F, \int E_\xi \varphi(\xi)\, d\xi \Big\rangle = \langle F, \hat{\varphi} \rangle = \langle \hat{F}, \varphi \rangle, \end{aligned}$$

so $G = \hat{F}$.

(8.51) Proposition. If $F \in \mathscr{E}'$, then $\hat{F}$ is a slowly increasing C^∞ function.

Proof. With notation as above, it is easy to check that $\partial^\alpha \hat{F}(\xi)$ exists in the classical sense and equals $\langle F, \partial_\xi^\alpha E_\xi \rangle = (-2\pi i)^{|\alpha|} \langle F, x^\alpha E_\xi \rangle$. Moreover, there exist C, N, m such that

$$|\partial^\alpha \hat{F}(\xi)| \le C \sum_{|\beta| \le N} \sup_{|x| \le m} \left| \partial_x^\beta \left(x^\alpha E_\xi(x) \right) \right| \le C'(1 + m)^{|\alpha|} (1 + |\xi|)^N,$$

so $\hat{F}$ is slowly increasing. □

This result yields an easy proof that every distribution is locally a distribution derivative of a continuous function.

(8.52) Proposition.

(a) If $F \in \mathscr{E}'$, there exist $N \in \mathbf{N}$, $f \in C_0(\mathbf{R}^n)$, and constants c_α ($|\alpha| \le N$) such that $F = \sum_{|\alpha| \le N} c_\alpha \partial^\alpha f$.

(b) If $F \in \mathscr{D}'$ and U is a precompact open set in $\mathbf{R}^n$, there exist N, f, c_α as above such that $F = \sum_{|\alpha| \le N} c_\alpha \partial^\alpha f$ on U.

Proof. By Proposition (8.51), if $F \in \mathscr{E}'$, there exists $M \in \mathbf{N}$ such that if $g(\xi) = (1 + 4\pi^2|\xi|^2)^{-M} \hat{F}(\xi)$, then $g \in L^1$. Let $f = \check{g}$: then $\hat{F} = (1 + 4\pi^2|\xi|^2)^M \hat{f}$, so $F = (I - \sum_1^n \partial_j^2)^M f$. This proves (a); for (b), choose $\psi \in C_c^\infty$ such that $\psi = 1$ on U, and apply (a) to ψF. □

EXERCISES

30. A locally integrable function f is in $\mathscr{S}'$ iff f is tempered.

31. If δ is the point mass at 0, then $(\partial^\alpha \delta)\hat{}\,(\xi) = (2\pi i \xi)^\alpha$ and $(x^\alpha)\hat{} = (-2\pi i)^{-|\alpha|} \partial^\alpha \delta$.

32. Let F be a distribution such that $\operatorname{supp}(F) = \{0\}$.

(a) There exist $N \in \mathbf{N}$, $C > 0$ such that for all $\varphi \in C_c^\infty$,

$$|\langle F, \varphi \rangle| \le C \sum_{|\alpha| \le N} \sup_{|x| \le 1} |\partial^\alpha \varphi(x)|.$$

(b) Fix $\psi \in C_c^\infty$ with $\psi(x) = 1$ for $|x| \le 1$ and $\psi(x) = 0$ for $|x| \ge 2$. If $\varphi \in C_c^\infty$, let $\varphi_k(x) = \varphi(x)[1 - \psi(kx)]$. If $\partial^\alpha \varphi(0) = 0$ for $|\alpha| \le N$, then $\partial^\alpha \varphi_k \to \partial^\alpha \varphi$ uniformly as $k \to \infty$ for $|\alpha| \le N$. [Hint: by Taylor's theorem, $|\partial^\alpha \varphi(x)| \le C|x|^{N+1-|\alpha|}$ for $|\alpha| \le N$.]

(c) If $\varphi \in C_c^\infty$ and $\partial^\alpha \varphi(0) = 0$ for $|\alpha| \le N$, then $\langle F, \varphi \rangle = 0$.

(d) There exist constants c_α ($|\alpha| \le N$) such that $\langle F, \varphi \rangle = \sum_{|\alpha| \le N} c_\alpha \partial^\alpha \varphi(0)$. In other words, if $b_\alpha = (-1)^{|\alpha|} c_\alpha$, $F = \sum_{|\alpha| \le N} b_\alpha \partial^\alpha \delta$ where δ is the point mass at 0.

33. A distribution F is called **harmonic** if $\sum_1^n \partial_j^2 F = 0$. [Every harmonic distribution is actually a C^∞ function, as we shall see in Theorem (8.65) below.] If F is harmonic and tempered, then F is a polynomial. (Consider $\hat{F}$ and use Exercises 31 and 32. This easily implies Liouville's theorem that every bounded harmonic function on $\mathbf{R}^n$ is constant.)

34. For $n \ge 3$, define F on $\mathbf{R}^n$ by

$$F(x) = \frac{|x|^{2-n}}{\omega_n(2 - n)} \quad \text{where } \omega_n = \frac{2\pi^{n/2}}{\Gamma(n/2)} = \text{volume of unit sphere.}$$

(a) F is a tempered function, hence a tempered distribution.

(b) Let $\Delta = \sum_1^n \partial_j^2$ and δ = point mass at 0; then $\Delta F = \delta$. [Let $G^\varepsilon(x) = (|x|^2 + \varepsilon^2)^{(2-n)/2}$. Then $G^\varepsilon \to |x|^{2-n}$ in $\mathscr{S}'$ as $\varepsilon \to 0$, so $\Delta G^\varepsilon \to \Delta(|x|^{2-n})$ in $\mathscr{S}'$. By explicit calculation, $\Delta G^\varepsilon(x) = g_\varepsilon(x) = \varepsilon^{-n} g(x/\varepsilon)$ where $g(x) = n(2-n)(|x|^2+1)^{-(n+2)/2}$. By Theorem (8.14), $g_\varepsilon \to [\int g]\delta$ as $\varepsilon \to 0$. To evaluate $\int g$, use polar coordinates and make the substitution $s = r^2/(r^2+1)$.]

(c) If $\varphi \in \mathscr{S}$, the function $f = \varphi * F$ satisfies the differential equation $\Delta f = \varphi$.

(d) These results hold also for $n = 1$ but can be proved much more simply there. For $n = 2$, they hold provided one takes $F(x) = (2\pi)^{-1}\log|x|$.

35. In this exercise we discuss the convolution of distributions.

(a) If $F \in \mathscr{E}'$, the operator $\varphi \to F * \varphi$ maps C_c^∞ into itself. By dualizing this map, define the convolution $F * G \in \mathscr{D}'$, for arbitrary $F \in \mathscr{E}'$ and $G \in \mathscr{D}'$. (We also set $G * F = F * G$ in this case.)

(b) If $F, G, H \in \mathscr{D}'$ and at most one of them has noncompact support, then $(F * G) * H = F * (G * H)$.

(c) On $\mathbf{R}^1$, let F be the constant function 1, $G = d\delta/dx$ where δ = point mass at 0, and $H = \chi_{(0,\infty)}$. Then $(F * G) * H = 0$ but $F * (G * H) = F$.

36. Let f be a continuous function on $\mathbf{R}^n \setminus \{0\}$ which is homogeneous of degree $-n$ [i.e., $f(rx) = r^{-n} f(x)$ for $r > 0$] and has mean zero on the unit sphere (i.e., $\int f\, d\sigma = 0$ where σ is surface measure on the sphere). Then f is not locally integrable near 0 unless $f = 0$, but f defines a tempered distribution $PV(f)$ (PV standing for "principal value") by

$$\langle PV(f), \varphi \rangle = \lim_{\varepsilon \to 0} \int_{|x|>\varepsilon} f(x)\varphi(x)\, dx.$$

(The limit equals

$$\int_{|x|\le 1} f(x)[\varphi(x) - \varphi(0)]\, dx + \int_{|x|>1} f(x)\varphi(x)\, dx,$$

and these integrals converge absolutely.)

37. On $\mathbf{R}^1$, let $F = PV((\pi x)^{-1})$ as in Exercise 36.

(a) $\hat{F}(\xi) = -i \operatorname{sgn} \xi$. [Let $F_{\varepsilon,N}(x) = (\pi x)^{-1}$ for $\varepsilon < |x| < N$, $F_{\varepsilon,N}(x) = 0$ otherwise. Then $F_{\varepsilon,N} \in L^1$; compute $\hat{F}_{\varepsilon,N}$ and use Exercise 2.61.]

(b) The map $\varphi \to F * \varphi$ ($\varphi \in \mathscr{S}$) extends to a unitary operator on L^2. (It is called the **Hilbert transform**.)

38. If $\lambda > n$ and $\varphi \in C_c^\infty$ satisfies $\varphi(0) > 0$, the value of the integral $\int \varphi(x)|x|^{-\lambda}\, dx$ is unequivocally $+\infty$. Nonetheless, there exist distributions which agree with $|x|^{-\lambda}$ on $\mathbf{R}^n \setminus \{0\}$. [If F is such a distribution, $\langle F, \varphi \rangle$ is sometimes called a **finite part** or a **renormalization** of the divergent integral $\int \varphi(x)|x|^{-\lambda}\, dx$.] Here are some procedures for constructing such distributions; the reader should verify that they work.

(a) If $\varphi \in C_c^\infty$, let P_φ^k be the Taylor polynomial of φ about $x = 0$ of degree k. If $k > \lambda - n - 1$ and $a > 0$, define

$$\langle F_a^k, \varphi \rangle = \int_{|x|\le a} \left[\varphi(x) - P_\varphi^k(x)\right]|x|^{-\lambda}\, dx + \int_{|x|>a} \varphi(x)|x|^{-\lambda}\, dx.$$

According to Exercise 32, $F_a^k - F_b^k$ is a linear combination of δ and its derivatives. Which one?

(b) If λ is not an integer and we choose k to satisfy $\lambda - n - 1 < k < \lambda - n$, we can let $a \to \infty$ in part (a) and define $\langle F, \varphi \rangle = \int [\varphi(x) - P_\varphi^k(x)]|x|^{-\lambda}\, dx$.

(c) For simplicity, we take $n = 1$ here. Let k be the greatest integer in λ; then $|x|^{-\lambda} = (d/dx)^k f(x)$ for $x \neq 0$, where

$$f(x) = \begin{cases} [(k-\lambda)\cdots(1-\lambda)]^{-1}(\operatorname{sgn} x)^k |x|^{k-\lambda} & \text{if } \lambda > k, \\ (-1)^{k-1}((k-1)!)^{-1}(\operatorname{sgn} x)^k \log|x| & \text{if } \lambda = k. \end{cases}$$

We have $f \in L^1_{loc}$, and the kth distribution derivative of f agrees with $|x|^{-\lambda}$ away from 0. Compare with (a) and (b).

39. (Distributions on $\mathbf{T}^n$) Give $C^\infty(\mathbf{T}^n)$ the Fréchet space topology defined by the seminorms $\|\varphi\|_{(\alpha)} = \|\partial^\alpha \varphi\|_u$. The space $\mathscr{D}'(\mathbf{T}^n)$ of distributions on $\mathbf{T}^n$ is the space of continuous linear functionals on $C^\infty(\mathbf{T}^n)$, with the weak * topology.

(a) Distributions on $\mathbf{T}^n$ can be translated, differentiated, and multiplied by C^∞ functions, just as on $\mathbf{R}^n$.

(b) If $F \in \mathscr{D}'(\mathbf{T}^n)$, its Fourier transform is the function $\hat{F}$ on $\mathbf{Z}^n$ defined by $\hat{F}(\kappa) = \langle F, E_\kappa \rangle$ where $E_\kappa(x) = e^{-2\pi i \kappa \cdot x}$. A function g on $\mathbf{Z}^n$ is the Fourier transform of a distribution on $\mathbf{T}^n$ iff $|g(\kappa)| \le C(1 + |\kappa|)^N$ for some $C, N > 0$. [If g is such a function, the series $\sum_\kappa g(\kappa) e^{2\pi i \kappa \cdot x}$ converges in $\mathscr{D}'(\mathbf{T}^n)$.]

(c) If $F \in \mathscr{D}'(\mathbf{T}^n)$ and $\varphi \in C^\infty(\mathbf{T}^n)$, then $\langle F, \bar{\varphi} \rangle = \sum_\kappa \hat{F}(\kappa)\overline{\hat{\varphi}(\kappa)}$.

(d) If $F, G \in \mathscr{D}'(\mathbf{T}^n)$ there is a unique $F * G \in \mathscr{D}'(\mathbf{T}^n)$ such that $(F * G)\hat{} = \hat{F}\hat{G}$. $F * G$ may also be defined by the two procedures in the text if $G \in C^\infty$ or by the procedure in Exercise 35 in the general case.

6. SOBOLEV SPACES

One of the most satisfactory ways of measuring smoothness properties of functions and distributions is in terms of L^2 norms. There are two reasons for

this: L^2 has the advantage of being a Hilbert space, and the Fourier transform, which converts differentiation into multiplication, is an isometry on L^2.

As a first step, suppose $k \in \mathbf{N}$ and let H_k be the space of all functions $f \in L^2(\mathbf{R}^n)$ whose distribution derivatives $\partial^\alpha f$ are L^2 functions for $|\alpha| \le k$. One can make H_k into a Hilbert space by defining a scalar product

$$\langle f, g \rangle = \sum_{|\alpha| \le k} \int (\partial^\alpha f)(\overline{\partial^\alpha g}).$$

However, it is more convenient to use an equivalent scalar product defined in terms of the Fourier transform. The Plancherel theorem implies that $f \in H_k$ iff $\xi^\alpha \hat{f} \in L^2$ for $|\alpha| \le k$. It is not hard to verify (Exercise 40) that there exist constants $C_1, C_2 > 0$ such that

$$C_1(1 + |\xi|^2)^k \le \sum_{|\alpha| \le k} |\xi^\alpha|^2 \le C_2(1 + |\xi|^2)^k,$$

from which it follows that $f \in H_k$ iff $(1 + |\xi|^2)^{k/2}\hat{f} \in L^2$ and that the norms

$$f \to \left(\sum_{|\alpha| \le k} \|\partial^\alpha f\|_2^2 \right)^{1/2} \quad \text{and} \quad f \to \left\| (1 + |\xi|^2)^{k/2} \hat{f} \right\|_2$$

are equivalent. The latter norm, however, makes sense for any $k \in \mathbf{R}$, and we can use it to extend the definition of H_k to all real k.

We proceed to the formal definitions. For any $s \in \mathbf{R}$ the function $(1 + |\xi|^2)^{s/2}$ is a slowly increasing C^∞ function, so the map Λ_s defined by

$$\Lambda_s f = \left((1 + |\xi|^2)^{s/2} \hat{f} \right)^\vee$$

is a continuous operator on $\mathcal{S}'$—actually an isomorphism, since $\Lambda_s^{-1} = \Lambda_{-s}$. If $s \in \mathbf{R}$, we define the **Sobolev space** H_s to be

$$H_s = \{ f \in \mathcal{S}' : \Lambda_s f \in L^2 \},$$

and we define an inner product and norm on H_s by

$$\langle f, g \rangle_{(s)} = \int (\Lambda_s f)(\overline{\Lambda_s g}), \qquad \|f\|_{(s)} = \|\Lambda_s f\|_2.$$

The following properties of Sobolev spaces are simple consequences of the definitions and the preceding discussion.

(i) The Fourier transform is a unitary isomorphism from H_s to $L^2(\mathbf{R}^n, (1 + |\xi|^2)^s \, d\xi)$. In particular, H_s is a Hilbert space.

(ii) $\mathcal{S}$ is a dense subspace of H_s for all $s \in \mathbf{R}$. [This follows easily from (i).]

(iii) If $s > t$, H_s is a dense subspace of H_t in the topology of H_t, and $\| \ \|_{(t)} \le \| \ \|_{(s)}$.

(iv) Λ_t is a unitary isomorphism from H_s to H_{s-t} for all s, t.

(v) $H_0 = L^2$ and $\| \ \|_{(0)} = \| \ \|_2$. By (ii), $H_s \subset L^2$ for $s > 0$. (For $s < 0$, the elements of H_s are in general not functions.)

(vi) ∂^α is a bounded linear map from H_s to $H_{s-|\alpha|}$ for all s, α, because $|\xi^\alpha| \le (1 + |\xi|^2)^{|\alpha|/2}$.

(8.53) Proposition. If $s \in \mathbf{R}$, the pairing between $\mathcal{S}'$ and $\mathcal{S}$ induces a unitary isomorphism from H_{-s} to $(H_s)^*$. More precisely, if $f \in H_{-s}$, the functional $\varphi \to \langle f, \varphi \rangle$ on $\mathcal{S}$ extends continuously to H_s with operator norm equal to $\|f\|_{(-s)}$, and every element of $(H_s)^*$ arises in this fashion.

Proof. If $f \in H_{-s}$ and $\varphi \in \mathcal{S}$,

$$\langle f, \varphi \rangle = \langle \check{f}, \hat{\varphi} \rangle = \int \check{f}(\xi)\hat{\varphi}(\xi) \, d\xi$$

since $\hat{f}$, and hence also $\check{f}$, is a tempered function. Thus by the Schwarz inequality,

$$|\langle f, \varphi \rangle| = \left[\int |\check{f}(\xi)|^2 (1 + |\xi|^2)^{-s} \, d\xi \right]^{1/2} \left[\int |\hat{\varphi}(\xi)|^2 (1 + |\xi|^2)^s \, d\xi \right]^{1/2}$$

$$= \|f\|_{(-s)} \|\varphi\|_{(s)},$$

so the functional $\varphi \to \langle f, \varphi \rangle$ extends to H_s, with norm at most $\|f\|_{(-s)}$. In fact, its norm equals $\|f\|_{(-s)}$, since if $g(x) = \Lambda_{-2s} f(-x)$, we have $g \in H_s$ and

$$\langle f, g \rangle = \int |\check{f}(\xi)|^2 (1 + |\xi|^2)^{-s} \, d\xi = \|f\|_{(-s)}^2 = \|f\|_{(-s)} \|g\|_{(s)}.$$

Finally, if $G \in (H_s)^*$, then $G \circ \mathcal{F}^{-1}$ is a bounded linear functional on $L^2((1 + |\xi|^2)^s d\xi)$, so there exists $g \in L^2((1 + |\xi|^2)^s d\xi)$ such that

$$G(\varphi) = \int (1 + |\xi|^2)^s g(\xi)\hat{\varphi}(\xi) \, d\xi.$$

But then $G(\varphi) = \langle f, \varphi \rangle$ where $f = \Lambda_{2s}\hat{g}$. (This argument would look a little neater if we took the pairing $\langle \ , \ \rangle$ between $\mathcal{S}$ and $\mathcal{S}'$ to be conjugate linear in the second variable, so that H_{-s} would become the conjugate dual of H_s.) □

For $s > 0$, the elements of H_s are functions which are "L^2-differentiable up to order s," and it is natural to ask what is the relationship between this notion

of smoothness and ordinary differentiability. This question has a simple and elegant answer. Let us introduce the notation

$$C_0^k = \{ f \in C^k(\mathbf{R}^n) \colon \partial^\alpha f \in C_0 \text{ for } |\alpha| \le k \}.$$

(8.54) The Sobolev Embedding Theorem. If $s > k + (n/2)$, then $H_s \subset C_0^k$.

Proof. Suppose $f \in H_s$; by the Fourier inversion theorem and the Riemann–Lebesgue lemma it will suffice to show that $(\partial^\alpha f)\hat{} \in L^1$ for $|\alpha| \le k$. But by the Schwarz inequality,

$$(2\pi)^{-|\alpha|}\int |(\partial^\alpha f)\hat{}(\xi)|\,d\xi = \int |\xi^\alpha \hat{f}(\xi)|\,d\xi \le \int (1+|\xi|^2)^{k/2}|\hat{f}(\xi)|\,d\xi$$

$$\le \left[\int (1+|\xi|^2)^s |\hat{f}(\xi)|^2\,d\xi\right]^{1/2}\left[\int (1+|\xi|^2)^{k-s}\,d\xi\right]^{1/2}.$$

The first factor on the right is $\|f\|_{(s)}$, and integration in polar coordinates shows that the second one is finite precisely when $2(k-s) < -n$, that is, when $s > k + (n/2)$. □

(8.55) Corollary. If $f \in H_s$ for all s, then $f \in C^\infty$.

An example may help to elucidate this theorem. Let $f_\lambda(x) = \varphi(x)|x|^\lambda$ where $\varphi \in C_c^\infty$ and $\varphi = 1$ on some neighborhood of 0. Then the (classical) derivative $\partial^\alpha f_\lambda$ is C^∞ except at 0 and is homogeneous of degree $\lambda - |\alpha|$ near 0, so that $|\partial^\alpha f_\lambda(x)| \le C_{\alpha,\lambda}|x|^{\lambda-|\alpha|}$, and in particular $\partial^\alpha f_\lambda \in L^1$ provided $\lambda - |\alpha| > -n$. In this case $\partial^\alpha f_\lambda$, as an L^1 function, is also the distribution derivative of f_λ. [To see this, replace f_λ by the C^∞ function $\varphi(x)(|x|^2+\varepsilon^2)^{\lambda/2}$ and consider the limit as $\varepsilon \to 0$.] Moreover, $\partial^\alpha f_\lambda \in L^2$ iff $\lambda - |\alpha| > -n/2$, so $f \in H_k$ $(k = 0, 1, 2, \dots)$ iff $\lambda > k - n/2$, whereas $f_\lambda \in C_0^k$ iff $\lambda > k$. See also Exercises 43–45 for some related results.

Next, we show that multiplication by suitably smooth functions preserves the H_s spaces. We need a lemma:

(8.56) Lemma. For all $\xi, \eta \in \mathbf{R}^n$ and $s \in \mathbf{R}$,

$$(1+|\xi|^2)^s(1+|\eta|^2)^{-s} \le 2^{|s|}(1+|\xi-\eta|^2)^{|s|}.$$

Proof. Since $|\xi| \le |\xi-\eta| + |\eta|$ we have $|\xi|^2 \le 2(|\xi-\eta|^2 + |\eta|^2)$ and hence

$$1 + |\xi|^2 \le 2(1+|\xi-\eta|^2)(1+|\eta|^2).$$

If $s \ge 0$, we have merely to raise both sides to the sth power. If $s < 0$, we

interchange ξ and η and replace s by $-s$, obtaining

$$(1+|\eta|^2)^{-s} \le 2^{-s}(1+|\xi|^2)^{-s}(1+|\xi-\eta|^2)^{-s},$$

which is again the desired result. □

(8.57) Theorem. Suppose $\varphi \in C_0(\mathbf{R}^n)$ and $\hat{\varphi}$ is a function which satisfies

$$\int (1+|\xi|^2)^{\alpha/2}|\hat{\varphi}(\xi)|\,d\xi = C < \infty \quad (\alpha > 0).$$

Then the map $M_\varphi \colon f \to \varphi f$ is a bounded operator on H_s for $|s| \le \alpha$.

Proof. Since Λ_s is a unitary map from H_s to $H_0 = L^2$, it is equivalent to show that $\Lambda_s M_\varphi \Lambda_{-s}$ is a bounded operator on L^2. But

$$(\Lambda_s M_\varphi \Lambda_{-s} f)\hat{}(\xi) = (1+|\xi|^2)^{s/2}[\hat{\varphi} * (\Lambda_{-s} f)\hat{}](\xi) = \int K(\xi,\eta)\hat{f}(\eta)\,d\eta$$

where

$$K(\xi,\eta) = (1+|\xi|^2)^{s/2}(1+|\eta|^2)^{-s/2}\hat{\varphi}(\xi-\eta).$$

By Lemma (8.56),

$$|K(\xi,\eta)| \le 2^{|s|/2}(1+|\xi-\eta|^2)^{|s|/2}|\hat{\varphi}(\xi-\eta)|,$$

so if $|s| \le \alpha$, $\int |K(\xi,\eta)|\,d\xi$ and $\int |K(\xi,\eta)|\,d\eta$ are bounded by $2^{|\alpha|/2}C$. That $\Lambda_s M_\varphi \Lambda_{-s}$ is bounded on L^2 therefore follows from the Plancherel theorem and Theorem (6.18). □

(8.58) Corollary. If $\varphi \in \mathscr{S}$, M_φ is a bounded operator on H_s for all s.

It is often useful to consider localized versions of the H_s spaces. If $\Omega \subset \mathbf{R}^n$ is open, a distribution f is said to be in $H_s^{loc}(\Omega)$ if for every precompact open U with $\overline{U} \subset \Omega$ there is a $g \in H_s$ which agrees with f on U.

(8.59) Proposition. A distribution f is in $H_s^{loc}(\Omega)$ iff $\varphi f \in H_s$ for all $\varphi \in C_c^\infty(\Omega)$.

Proof. If $f \in H_s^{loc}(\Omega)$ and $\varphi \in C_c^\infty(\Omega)$, then f agrees with some $g \in H_s$ on a neighborhood of $\operatorname{supp}(\varphi)$; hence $\varphi f = \varphi g \in H_s$ by Corollary (8.58). For the converse, if U is precompact and open and $\overline{U} \subset \Omega$, we can choose $\varphi \in C_c^\infty(\Omega)$ with $\varphi = 1$ on $\overline{U}$ by the C^∞ Urysohn lemma; then $\varphi f \in H_s$ and $\varphi f = f$ on U. □

Our final result in this section is a compactness theorem which is of great importance in the applications of Sobolev spaces.

(8.60) Rellich's Theorem. Suppose $\{f_k\}$ is a sequence in H_s such that (i) $\sup_k \|f\|_{(s)} < \infty$, and (ii) the f_k's are all supported in a fixed compact set V. Then there is a subsequence $\{f_{k_j}\}$ which converges in H_t for all $t < s$.

Proof. First we observe that by Proposition (8.51), $\hat{f}_k$ is a slowly increasing C^∞ function. Pick $\varphi \in C_c^\infty$ such that $\varphi = 1$ on V. Then $f_k = \varphi f_k$, so $\hat{f}_k = \hat{\varphi} * \hat{f}_k$ where the convolution is defined pointwise by an absolutely convergent integral. By Lemma (8.56) and the Schwarz inequality,

$$(1 + |\xi|^2)^{s/2} |\hat{f}_k(\xi)| \le 2^{|s|/2} \int |\hat{\varphi}(\xi - \eta)| (1 + |\xi - \eta|^2)^{|s|/2}$$

$$\times |\hat{f}_k(\eta)| (1 + |\eta|^2)^{s/2} \, d\eta$$

$$\le 2^{|s|/2} \|\varphi\|_{(|s|)} \|f_k\|_{(s)} \le \text{constant}.$$

Likewise, since $\partial_j(\hat{\varphi} * \hat{f}_k) = (\partial_j \hat{\varphi}) * \hat{f}_k$, $(1 + |\xi|^2)^{s/2} |\partial_j \hat{f}_k(\xi)|$ is bounded by a constant independent of ξ, j, and k. In particular, the $\hat{f}_k$'s and their first derivatives are uniformly bounded on compact sets, so by the mean value theorem and the Ascoli–Arzelà theorem there is a subsequence $\{\hat{f}_{k_j}\}$ which converges uniformly on compact sets.

We claim that $\{f_{k_j}\}$ is Cauchy in H_t for all $t < s$. Indeed, for any $R > 0$ we can write the integral

$$\|f_{k_i} - f_{k_j}\|_{(t)}^2 = \int (1 + |\xi|^2)^t |\hat{f}_{k_i} - \hat{f}_{k_j}|^2(\xi) \, d\xi$$

as the sum of the integrals over the regions $|\xi| \le R$ and $|\xi| > R$. For $|\xi| \le R$ we use the estimate

$$(1 + |\xi|^2)^t \le (1 + R^2)^T \quad [T = \max(t, 0)],$$

and for $|\xi| > R$ we use the estimate

$$(1 + |\xi|^2)^t \le (1 + R^2)^{t-s} (1 + |\xi|^2)^s,$$

which yield

$$\|f_{k_i} - f_{k_j}\|_{(t)}^2 \le CR^n (1 + R^2)^T \sup_{|\xi| \le R} |\hat{f}_{k_i} - \hat{f}_{k_j}|^2(\xi)$$

$$+ (1 + R^2)^{t-s} \|f_{k_i} - f_{k_j}\|_{(s)}^2.$$

Given $\varepsilon > 0$, the second term will be less than $\varepsilon/2$ provided R is chosen sufficiently large (since $t - s < 0$); and having fixed such an R, we can make the first term less than $\varepsilon/2$ by choosing i and j sufficiently large. The proof is therefore complete. □

EXERCISES

40. Suppose $k \in \mathbf{N}$.

(a) There exist $A_1, A_2 > 0$ such that $A_1|\xi|^{2k} \le \sum_{|\alpha|=k} |\xi^\alpha|^2 \le A_2|\xi|^{2k}$ for all $\xi \in \mathbf{R}^n$. (The quantities in question are homogeneous of degree $2k$, hence are determined by their values on the unit sphere.)

(b) There exist $C_1, C_2 > 0$ such that for all $\xi \in \mathbf{R}^n$,

$$C_1(1 + |\xi|^2)^k \le \sum_{|\alpha| \le k} |\xi^\alpha|^2 \le C_2(1 + |\xi|^2)^k.$$

41. If $k \in \mathbf{N}$, H_k is the space of all L^2 functions whose L^2 derivatives up to order k (in the sense of Exercise 8) exist, and these strong derivatives coincide with the distribution derivatives.

42. Suppose $r < s < t$. For any $\varepsilon > 0$ there exists $C > 0$ such that $\|f\|_{(s)} \le \varepsilon \|f\|_{(t)} + C\|f\|_{(r)}$ for all $f \in H_t$.

43. (Converse of the Sobolev theorem) If $H_s \subset C_0^k$, then $s > k + (n/2)$. [C_0^k is a Banach space in a natural way. If $H_s \subset C_0^k$, then the inclusion map, and hence the evaluation functionals $f \to \partial^\alpha f(0)$, are continuous. Use Exercise 31 and Proposition (8.53).]

44. (A sharper Sobolev theorem) For $0 < \alpha < 1$, let

$$\Lambda_\alpha(\mathbf{R}^n) = \{ f \in BC(\mathbf{R}^n) : \sup_{x \ne y} |f(x) - f(y)|/|x - y|^\alpha < \infty \}.$$

(a) If $s = (n/2) + \alpha$ where $0 < \alpha < 1$, then $\|\delta_x - \delta_y\|_{(-s)} \le C_\alpha |x - y|^\alpha$ where δ_x, δ_y are the point masses at x and y. [We have $\hat{\delta}_x(\xi) = e^{-2\pi i x \cdot \xi}$. Write the integral defining $\|\delta_x - \delta_y\|_{(-s)}^2$ as the sum of the integrals over the regions $|\xi| \le R$ and $|\xi| > R$ where $R = |x - y|^{-1}$, and use the mean value theorem to estimate $\hat{\delta}_x - \hat{\delta}_y$ on the first region.]

(b) If $s = (n/2) + \alpha$ where $0 < \alpha < 1$, then $H_s \subset \Lambda_\alpha$.

(c) If $s = (n/2) + k + \alpha$ where $k \in \mathbf{N}$ and $0 < \alpha < 1$, then $H_s \subset \{ f \in C_0^k : \partial^\alpha f \in \Lambda_\alpha \text{ for } |\alpha| = k \}$.

45. The Sobolev theorem says that if $s > n/2$ it makes sense to evaluate elements of H_s at a point. For $0 \le s < n/2$, functions in H_s are only defined a.e., but if $s > k/2$ it makes sense to restrict functions in H_s to subspaces of codimension k. More precisely, let us write $\mathbf{R}^n = \mathbf{R}^{n-k} \times \mathbf{R}^k$,

$x = (y, z)$, $\xi = (\eta, \zeta)$ and define R: $\mathscr{S}(\mathbf{R}^n) \to \mathscr{S}(\mathbf{R}^{n-k})$ by $Rf(y) = f(y, 0)$.

(a) $(Rf)\hat{}(\eta) = \int \hat{f}(\eta, \zeta)\, d\zeta$.

(b) If $s > k/2$,

$$|(Rf)\hat{}(\eta)|^2 \le C_s(1 + |\eta|^2)^{(k/2)-s} \int |\hat{f}(\eta, \zeta)|^2 (1 + |\eta|^2 + |\zeta|^2)^s \, d\zeta.$$

(c) R extends to a bounded map from $H_s(\mathbf{R}^n)$ to $H_{s-(k/2)}(\mathbf{R}^{n-k})$ provided $s > k/2$.

46. Suppose $s_0 \le s_1$, $t_0 \le t_1$. If T is a bounded linear map from H_{s_0} to H_{t_0} whose restriction to H_{s_1} is bounded from H_{s_1} to H_{t_1}, then the restriction of T to H_{s_λ} is bounded from H_{s_λ} to H_{t_λ} for $0 < \lambda < 1$, where $s_\lambda = (1 - \lambda)s_0 + \lambda s_1$ and $t_\lambda = (1 - \lambda)t_0 + \lambda t_1$. [$T$ is bounded from H_s to H_t iff $\Lambda_t T \Lambda_{-s}$ is bounded on L^2. Observe that Λ_z is well defined for all $z \in \mathbf{C}$ and Λ_z is unitary on every H_s if $\operatorname{Re} z = 0$. Let $s(z) = (1 - z)s_0 + zs_1$, $t(z) = (1 - z)t_0 + zt_1$, and for $0 \le \operatorname{Re} z \le 1$ and $\varphi, \psi \in \mathscr{S}$ let $F(z) = \int [\Lambda_{t(z)} T \Lambda_{-s(z)} \varphi]\bar{\psi}$. Apply the three lines lemma as in the proof of the Riesz–Thorin theorem.]

47. Let $\Omega \subset \mathbf{R}^n$ be open and let F: $\Omega \to \mathbf{R}^n$ be a C^∞ diffeomorphism. Then the operator $f \to f \circ F$ maps $H_s^{loc}(F(\Omega))$ into $H_s^{loc}(\Omega)$ for all s. More precisely, for any $\varphi \in C_c^\infty(F(\Omega))$, the operator $Tf = (\varphi f) \circ F$ is bounded on H_s. Proceed as follows:

(a) If $s = 0, 1, 2, \ldots$, use the chain rule and the fact that $H_s = \{f: \partial^\alpha f \in L^2 \text{ for } |\alpha| \le s\}$.

(b) Use Exercise 46 to obtain the result for all $s \ge 0$.

(c) For $s < 0$, use Proposition (8.53) and the fact that the adjoint of T is another operator of the same type, namely $T^* f = (\psi f) \circ G$ where $G = F^{-1}$, $\psi = (J\varphi) \circ F$, and $J(x) = |\det D_x G|$.

48. Recall the discussion of distributions on $\mathbf{T}^n$ in Exercise 39, and define

$$H_s(\mathbf{T}^n) = \left\{ f \in \mathscr{D}'(\mathbf{T}^n): \sum (1 + |\kappa|^2)^s |\hat{f}(\kappa)|^2 < \infty \right\}.$$

State and prove analogs of the results in this section for $H_s(\mathbf{T}^n)$.

7. APPLICATIONS TO PARTIAL DIFFERENTIAL EQUATIONS

In this section we present a few of the many applications of Fourier analysis to the theory of partial differential equations. We shall use the term **differential operator** to mean a linear partial differential operator with smooth coefficients, that is, an operator of the form

$$Lf = \sum_{|\alpha| \le m} a_\alpha \partial^\alpha f, \quad a_\alpha \in C^\infty.$$

If the a_α's are constants, we call L a **constant-coefficient** operator. In this case, we have

$$(Lf)\hat{}(\xi) = \sum_{|\alpha| \le m} a_\alpha (2\pi i \xi)^\alpha \hat{f}(\xi).$$

It is therefore convenient to write L in a slightly different form: we replace a_α by $(2\pi i)^{|\alpha|} a_\alpha$ and introduce the operators

$$D^\alpha = (2\pi i)^{-|\alpha|} \partial^\alpha,$$

so that $L = \sum_{|\alpha| \le m} a_\alpha D^\alpha$ and $(Lf)\hat{} = \sum_{|\alpha| \le m} a_\alpha \xi^\alpha \hat{f}$. Thus if P is any polynomial in n variables over $\mathbf{C}$, say $P(\xi) = \sum_{|\alpha| \le m} a_\alpha \xi^\alpha$, we can form the constant-coefficient operator $P(D) = \sum_{|\alpha| \le m} a_\alpha D^\alpha$, and we then have $(P(D)f)\hat{} = P\hat{f}$. The polynomial P is called the **symbol** of the operator $P(D)$, and its top-order part

$$P_m(\xi) = \sum_{|\alpha| = m} a_\alpha \xi^\alpha$$

is called the **principal symbol** of $P(D)$.

The most important of all differential operators is the **Laplacian**

$$\Delta = \sum_1^n \partial_j^2 = -4\pi^2 \sum_1^n D_j^2.$$

The reason for this is that Δ is essentially the only (scalar) differential operator which is invariant under translations and rotations. (If one considers operators on vector-valued functions, there are others, such as the familiar div, curl, and grad of three-dimensional vector analysis.) More precisely, we have:

(8.61) Theorem. A differential operator L satisfies $L(f \circ T) = (Lf) \circ T$ for all translations and rotations T iff there is a polynomial P in one variable such that $L = P(\Delta)$.

Proof. Clearly L is translation-invariant iff L has constant coefficients, in which case $L = Q(D)$ for some polynomial Q in n variables. Moreover, since $(Lf)\hat{} = Q\hat{f}$ and the Fourier transform commutes with rotations, L is rotation-invariant iff Q is rotation-invariant. Let $Q = \sum_0^m Q_j$ where Q_j is homogeneous of degree j: then it is easy to see that Q is rotation-invariant iff each Q_j is rotation-invariant. [Use induction on j and the fact that $Q_j(\xi) = \lim_{r \to 0} r^{-j} \sum_j^m Q_i(r\xi)$.] But this means that $Q_j(\xi)$ depends only on $|\xi|$, and homogeneity then implies that $Q_j(\xi) = c_j |\xi|^j$. Moreover, $|\xi|^j$ is a polynomial precisely when j is even, so $c_j = 0$ for j odd, and if $b_k = (-4\pi^2)^{-k} c_{2k}$, we have $Q(\xi) = \sum b_k (-4\pi^2 |\xi|^2)^k$, that is, $L = \sum b_k \Delta^k$. □

One of the basic boundary value problems for the Laplacian is the **Dirichlet problem**: given an open set $\Omega \subset \mathbf{R}^n$ and a function f on its boundary $\partial\Omega$, find a function u on $\overline{\Omega}$ such that $\Delta u = 0$ on Ω and $u|\partial\Omega = f$. (This statement of the problem is deliberately a bit imprecise.) We shall solve the Dirichlet problem when Ω is a half-space. For this purpose it will be convenient to replace n by $n+1$, so that the half-space will be $\mathbf{R}^n \times (0, \infty)$, and to denote the coordinates on $\mathbf{R}^{n+1}$ by $x_1, \ldots, x_n, t$. We continue to use the symbol Δ to denote the Laplacian on $\mathbf{R}^n$, so the Laplacian on $\mathbf{R}^{n+1}$ is $\Delta + \partial_t^2$ where $\partial_t = \partial/\partial t$. Given a function f on $\mathbf{R}^n$, satisfying conditions to be made more precise below, we wish to find a function u on $\mathbf{R}^n \times [0, \infty)$ such that $(\Delta + \partial_t^2)u = 0$ and $u(x, 0) = f(x)$.

The idea is to apply the Fourier transform on $\mathbf{R}^n$, thus converting the partial differential equation $(\Delta + \partial_t^2)u = 0$ into the ordinary differential equation $(-4\pi^2|\xi|^2 + \partial_t^2)\hat{u} = 0$. The general solution of this equation is

$$\hat{u}(\xi, t) = c_1(\xi)e^{-2\pi|\xi|t} + c_2(\xi)e^{2\pi|\xi|t}.$$

We reject the second term (i.e., take $c_2 = 0$) because $e^{2\pi|\xi|t}$ is not tempered and so does not have an inverse Fourier transform. Setting $t = 0$, we have $c_1(\xi) = \hat{u}(\xi, 0) = \hat{f}(\xi)$, so $\hat{u}(\xi, t) = \hat{f}(\xi)e^{-2\pi|\xi|t}$. Therefore, $u(x, t) = (f * P_t)(x)$ where $P_t = (e^{-2\pi|\xi|t})\hat{}$. But we have calculated this in Exercise 13: P_t is the Poisson kernel,

$$P_t(x) = \frac{\Gamma((n+1)/2)}{\pi^{(n+1)/2}} \frac{t}{(t^2 + |x|^2)^{(n+1)/2}}.$$

So far this is all formal, since we have not specified conditions on f to ensure that these manipulations are justified. We now give a precise result.

(8.62) Theorem. Suppose $f \in L^p(\mathbf{R}^n)$, $1 \le p \le \infty$. Then the function $u(x, t) = (f * P_t)(x)$ satisfies $(\Delta + \partial_t^2)u = 0$ on $\mathbf{R}^n \times (0, \infty)$; moreover, $\lim_{t \to 0} u(x, t) = f(x)$ for a.e. x (and for every x at which f is continuous), and $\lim_{t \to 0} \|u(\cdot, t) - f\|_p = 0$ provided $p < \infty$.

Proof. P_t and all of its derivatives are in $L^q(\mathbf{R}^n)$ for $1 \le q \le \infty$, since a rough calculation shows that $|\partial_x^\alpha P_t(x)| \le C_t|x|^{-n-1-|\alpha|}$ and $|\partial_t^j P_t(x)| \le C_t|x|^{-n-1}$ for large x. Also $(\Delta + \partial_t^2)P_t = 0$, which can be verified by direct computation or (more easily) by taking the Fourier transform. Hence $f * P_t$ is well defined and

$$(\Delta + \partial_t^2)(f * P_t) = f * (\Delta + \partial_t^2)P_t = 0.$$

Since $P_t(x) = t^{-n}P_1(x/t)$ and $\int P_1(x)\,dx = \hat{P}_1(0) = 1$, the remaining assertions follow from Theorems (8.14) and (8.15). □

The same idea can be used to solve the **heat equation** $(\partial_t - \Delta)u = 0$ on $\mathbf{R}^n \times (0, \infty)$ subject to the initial condition $u(x, 0) = f(x)$. [Physical interpretation: $u(x, t)$ represents the temperature at position x and time t in a homogeneous isotropic medium, given that the temperature at time 0 is $f(x)$.] Indeed, Fourier transformation leads to the ordinary differential equation $(\partial_t + 4\pi^2|\xi|^2)\hat{u} = 0$ with initial condition $\hat{u}(\xi, 0) = \hat{f}(\xi)$. The unique solution of this problem is $\hat{u}(\xi, t) = \hat{f}(\xi)e^{-4\pi^2|\xi|^2 t}$, and hence, by Proposition (8.24), $u(x, t) = f * G_t(x)$ where

$$G_t(x) = (4\pi t)^{-n/2}e^{-|x|^2/4t}.$$

Here we have $G_t(x) = t^{-n/2}G_1(x/\sqrt{t})$, so after the change of variable $s = \sqrt{t}$, Theorems (8.14) and (8.15) apply again, and we obtain an exact analog of Theorem (8.62) for the initial value problem $(\partial_t - \Delta)u = 0$, $u(x, 0) = f(x)$. Actually, in the present case we can relax the hypotheses on f considerably because $G_t \in \mathcal{S}$: see Exercise 49.

Another fundamental differential equation of mathematical physics is the **wave equation** $(\partial_t^2 - \Delta)u = 0$. [Physical interpretation: $u(x, t)$ is the amplitude at position x and time t of a wave traveling in a homogeneous isotropic medium with speed 1.] Here it is appropriate to impose two initial conditions at $t = 0$: $u(x, 0) = f(x)$ and $\partial_t u(x, 0) = g(x)$. After applying the Fourier transform, we obtain

$$(\partial_t^2 + 4\pi^2|\xi|^2)\hat{u}(\xi, t) = 0, \quad \hat{u}(\xi, 0) = \hat{f}(\xi), \quad \partial_t\hat{u}(\xi, 0) = \hat{g}(\xi),$$

the solution to which is

$$\hat{u}(\xi, t) = \cos(2\pi|\xi|t)\hat{f}(\xi) + \frac{\sin(2\pi|\xi|t)}{2\pi|\xi|}\hat{g}(\xi).$$

Since

$$\cos(2\pi|\xi|t) = \frac{\partial}{\partial t}\left[\frac{\sin(2\pi|\xi|t)}{2\pi|\xi|}\right],$$

it follows that

$$u(x, t) = f * \partial_t W_t(x) + g * W_t(x), \quad \text{where } W_t = \left[\frac{\sin(2\pi|\xi|t)}{2\pi|\xi|}\right]\check{}.$$

Unfortunately, evaluating W_t is not so simple, for W_t is a distribution which becomes increasingly singular as the space dimension n increases. [Indeed, it is clear that $W_t \in H_s$ iff $s < 1 - (n/2)$.] For $n \le 3$, however, the calculation can be carried out without undue difficulty: see Exercise 50.

If $L = \sum_0^m a_j(d/dx)^j$ is an ordinary differential operator with C^∞ coefficients such that a_m never vanishes, it is not hard to show that smooth data give smooth solutions. More precisely, if $Lu = f$ and f is C^k on an open interval I, then u is C^{k+m} on I. No such result holds for partial differential operators in

general. For example, the distribution W_t discussed above satisfies $(\partial_t^2 - \Delta)W_t = 0$ for $t > 0$, but W_t is not even continuous, much less C^∞. However, there is a large class of differential operators for which a strong regularity theorem holds. We restrict attention to the constant-coefficient case, although the results are valid in greater generality.

A constant-coefficient operator $P(D)$ of order m is called **elliptic** if its principal symbol $P_m(\xi)$ is nonzero for $\xi \neq 0$. Thus, ellipticity means that, in a formal sense, $P(D)$ is genuinely mth order in all directions. (For example, Δ is elliptic on $\mathbf{R}^n$, whereas $\partial_t - \Delta$ and $\partial_t^2 - \Delta$ are not elliptic on $\mathbf{R}^{n+1}$.) The following characterization of ellipticity will be useful.

(8.63) Lemma. Suppose $P(D)$ is a differential operator of order m. Then $P(D)$ is elliptic iff there exist $C, R > 0$ such that $|P(\xi)| \geq C|\xi|^m$ when $|\xi| \geq R$.

Proof. If $P(D)$ is elliptic, let C_1 be the minimum value of $P_m(\xi)$ on the unit sphere $|\xi| = 1$. Then $C_1 > 0$, and since P_m is homogeneous of degree m we have $|P_m(\xi)| \geq C_1|\xi|^m$ for all ξ. On the other hand, $P - P_m$ is of order $m - 1$, so there exists C_2 such that $|P(\xi) - P_m(\xi)| \leq C_2|\xi|^{m-1}$. Therefore,

$$|P(\xi)| \geq |P_m(\xi)| - |P(\xi) - P_m(\xi)| \geq \tfrac{1}{2}C_1|\xi|^m \text{ for } |\xi| \geq 2C_2C_1^{-1}.$$

Conversely, if $P(D)$ is not elliptic, say $P_m(\xi_0) = 0$, then $|P(\xi)| \leq C|\xi|^{m-1}$ for all ξ on the line spanned by ξ_0. □

(8.64) Lemma. If $P(D)$ is elliptic of order m, $u \in H_s$, and $P(D)u \in H_s$, then $u \in H_{s+m}$.

Proof. The hypotheses say that $(1 + |\xi|^2)^{s/2}\hat{u} \in L^2$ and $(1 + |\xi|^2)^{s/2}P\hat{u} \in L^2$. By Lemma (8.63), for some $R \geq 1$ we have

$$(1 + |\xi|^2)^{m/2} \leq 2^m|\xi|^m \leq C^{-1}2^m|P_m(\xi)| \text{ for } |\xi| \geq R,$$

and $(1 + |\xi|^2)^{m/2} \leq (1 + R^2)^{m/2}$ for $|\xi| \leq R$. It follows that

$$(1 + |\xi|^2)^{(s+m)/2}|\hat{u}| \leq C_1(1 + |\xi|^2)^{s/2}[|P\hat{u}| + |\hat{u}|] \in L^2,$$

that is, $u \in H_{s+m}$. □

(8.65) The Elliptic Regularity Theorem. Suppose L is a constant-coefficient elliptic differential operator of order m, Ω is open in $\mathbf{R}^n$, and $u \in \mathscr{D}'$. If $Lu \in H_s^{loc}(\Omega)$ for some $s \in \mathbf{R}$ then $u \in H_{s+m}^{loc}(\Omega)$, and if $L_u \in C^\infty(\Omega)$ then $u \in C^\infty(\Omega)$.

Proof. The second assertion follows from the first in view of Corollary (8.55), so by Proposition (8.59) we must show that if $Lu \in H_s^{loc}(\Omega)$ and

$\varphi \in C_c^\infty(\Omega)$, then $\varphi u \in H_{s+m}$. Let U be a precompact open set such that $\text{supp}(\varphi) \subset U \subset \overline{U} \subset \Omega$, and choose $\psi \in C_c^\infty(\Omega)$ such that $\psi = 1$ on $\overline{U}$. Then $\psi u \in \mathscr{E}'$, so it follows from Proposition (8.51) that $\psi u \in H_\sigma$ for some $\sigma \in \mathbf{R}$; by decreasing σ we may assume that $s + m - \sigma$ is a positive integer k. Set $\psi_0 = \psi$ and $\psi_k = \varphi$, and choose recursively $\psi_1, \ldots, \psi_{k-1} \in C_c^\infty$ such that $\psi_j = 1$ on a neighborhood of $\text{supp}(\varphi)$ and $\text{supp}(\psi_j) \subset \{x\colon \psi_{j-1}(x) = 1\}$. We shall prove by induction that $\psi_j u \in H_{\sigma+j}$. When $j = k$ we obtain $\varphi u = \psi_k u \in H_{\sigma+k} = H_{s+m}$, which will complete the proof.

The crucial observation is that for any $\zeta \in C_c^\infty$ the operator $[L, \zeta]$ defined by

$$[L, \zeta]f = L(\zeta f) - \zeta Lf$$

is a differential operator of order $m - 1$ whose coefficients are derivatives of ζ; in particular, they are C^∞ functions supported in $\text{supp}(\zeta)$ which vanish on any open set where ζ is constant. (This follows from the product rule for derivatives.) Thus if $f \in H_t$ we have $[L, \zeta]f \in H_{t-(m-1)}$ by Theorem (8.57).

For $j = 0$ we have $\psi_0 u \in H_\sigma$ by assumption. Suppose we have established that $\psi_j u \in H_{\sigma+j}$, where $0 \leq j < k$. Then by the preceding remarks,

$$L(\psi_{j+1}u) = \psi_{j+1}Lu + [L, \psi_{j+1}]u = \psi_{j+1}Lu + [L, \psi_{j+1}]\psi_j u$$

$$\in H_s + H_{\sigma+j-m+1} = H_{\sigma+j-m+1}.$$

Since $\psi_{j+1}u = \psi_{j+1}\psi_j u \in H_{\sigma+j}$, Lemma (8.64) implies that $\psi_{j+1}u \in H_{\sigma+j+1}$, and we are done. □

Two classical special cases of this theorem are particularly noteworthy. First, every distribution solution of Laplace's equation $\Delta u = 0$ is a C^∞ function. (This fact is known as **Weyl's lemma**.) Second, if $L = \partial_1 + i\partial_2$ on $\mathbf{R}^2$, the equation $Lu = 0$ is the **Cauchy–Riemann equation**, whose solutions are the holomorphic (or analytic) functions of $z = x_1 + ix_2$. We thus recover the fact that holomorphic functions are C^∞.

EXERCISES

49. Let $G_t(x) = (4\pi t)^{-n/2}e^{-|x|^2/4t}$, and if $f \in \mathscr{S}'(\mathbf{R}^n)$, let $u(x, t) = f * G_t(x)$.

(a) u satisfies $(\partial_t - \Delta)u = 0$ on $\mathbf{R}^n \times (0, \infty)$ and $u(\cdot, t) \to f$ in $\mathscr{S}'$ as $t \to 0$.

(b) If f is a tempered function, then $u(x, t) \to f(x)$ a.e. as $t \to 0$. [To show that $u(\cdot, t) \to f$ a.e. on a bounded open set V, write $f = \varphi f + (1 - \varphi)f$ where $\varphi \in C_c^\infty$ and $\varphi = 1$ on V, and show that $((1 - \varphi)f) * G_t \to 0$ on V.] Can you relax the growth hypotheses on f still further?

50. Let W_t be the inverse Fourier transform of $(2\pi|\xi|)^{-1}\sin(2\pi|\xi|t)$.

(a) If $n = 1$, $W_t = \frac{1}{2}\chi_{(-t,t)}$. (Use the Fourier inversion theorem.)

(b) If $n = 3$, let σ_R denote surface measure on the sphere $|x| = R$. Then $\hat{\sigma}_R(\xi) = 2R|\xi|^{-1}\sin(2\pi R|\xi|)$, and hence $W_t = (4\pi t)^{-1}\sigma_t$. (This calculation is essentially contained in Exercise 19d.)

(c) If $n = 2$, think of $\xi \in \mathbf{R}^2$ as an element of $\mathbf{R}^2 \times \{0\} \subset \mathbf{R}^3$. Transform the integral

$$\frac{2R\sin(2\pi R|\xi|)}{|\xi|} = \int_{|x|=R} e^{-2\pi i x\cdot\xi}\, d\sigma_R(x)$$

in part (b) into an integral over the disc $D_R = \{y\colon |y| \le R\}$ in $\mathbf{R}^2$ by projecting the upper and lower hemispheres of the sphere in $\mathbf{R}^3$ onto the equatorial plane. Conclude that for $n = 2$,

$$W_t(x) = (2\pi)^{-1}\left(t^2 - |x|^2\right)^{-1/2}\chi_{D_t}(x).$$

(d) Write out $u = f * \partial_t W_t + g * W_t$ explicitly for $n = 1, 2, 3$. Observe that $\operatorname{supp}(W_t) \subset \{x\colon |x| \le t\}$, which expresses the fact that the waves propagate with speed 1. In particular, $W_t \in \mathscr{E}'$, so by Exercise 35 the formula for u makes sense for any $f, g \in \mathscr{D}'$, and u satisfies the wave equation in the sense of distributions.

51. Work out the analogs of the solutions of the boundary value problems discussed in the text when $\mathbf{R}^n$ is replaced by $\mathbf{T}^n$.

52. In this exercise we discuss heat flow on an interval.

(a) Solve $(\partial_t - \partial_x^2)u = 0$ on $(a, b) \times (0, \infty)$ with boundary conditions $u(x, 0) = f(x)$ on (a, b), $u(a, t) = u(b, t) = 0$ for $t > 0$. [This describes heat flow on (a, b) when the endpoints are held at a constant temperature. It suffices to assume $a = 0$, $b = \frac{1}{2}$; extend f to $\mathbf{R}$ by requiring f to be odd and periodic, and use Exercise 51.]

(b) Same problem, with the boundary condition $u(a, t) = u(b, t) = 0$ replaced by $\partial_x u(a, t) = \partial_x u(b, t) = 0$. [This describes heat flow on (a, b) when the endpoints are insulated. This time, extend f to be even and periodic.]

53. Solve the vibrating string problem: $(\partial_t^2 - \partial_x^2)u = 0$ on $(a, b) \times (0, \infty)$, $u(x, 0) = f(x)$ and $\partial_t u(x, 0) = g(x)$ on (a, b), $u(a, t) = u(b, t) = 0$ for $t > 0$. (Cf. Exercise 52.)

54. Let $G(x, t) = G_t(x)$ as in Exercise 49 for $t > 0$, and $G(x, t) = 0$ for $t \le 0$.

(a) G is a tempered function on $\mathbf{R}^{n+1}$. If $G^\varepsilon(x, t) = G(x, t)$ for $t > \varepsilon$ and $G^\varepsilon(x, t) = 0$ for $t \le \varepsilon$, then $G^\varepsilon \to G$ in $\mathscr{S}'$ as $\varepsilon \to 0$.

(b) If $f \in \mathscr{S}(\mathbf{R}^{n+1})$ and $u = f * G$ (convolution on $\mathbf{R}^{n+1}$), then $(\partial_t - \Delta)u = f$. (Hint: $u = \lim_{\varepsilon\to 0} f * G^\varepsilon$.)

55. The heat operator $\partial_t - \Delta$ is not elliptic, but a weakened version of Theorem (8.65) holds for it. Here we are working on $\mathbf{R}^{n+1}$ with coordinates (x, t) and dual coordinates (ξ, τ), and $\partial_t - \Delta = P(D_x, D_t)$ where $P(\xi, \tau) = 2\pi i\tau + 4\pi^2|\xi|^2$.

(a) There exist $C, R > 0$ such that $|\xi|\,|(\xi, \tau)|^{1/2} \le C|P(\xi, \tau)|$ for $|(\xi, \tau)| > R$. (Consider the regions $|\tau| \le |\xi|^2$ and $|\tau| \ge |\xi|^2$ separately.)

(b) If $f \in H_s$ and $(\partial_t - \Delta)f \in H_s$, then $f \in H_{s+1}$ and $(\partial f/\partial x_i) \in H_{s+(1/2)}$ for $1 \le i \le n$.

(c) If $\zeta \in C_c^\infty(\mathbf{R}^{n+1})$ we have

$$[\partial_t - \Delta, \zeta]f = (\partial_t\zeta - \Delta\zeta)f - 2\sum_1^n \left(\frac{\partial\zeta}{\partial x_i}\right)\left(\frac{\partial f}{\partial x_i}\right).$$

(d) If $u \in \mathscr{D}'$, Ω is open in $\mathbf{R}^{n+1}$, and $(\partial_t - \Delta)u \in H_s^{loc}(\Omega)$, then $u \in H_{s+1}^{loc}(\Omega)$. [Let ψ_j be as in the proof of Theorem (8.65). Show inductively that if $\psi_0 u \in H_\sigma$, then $\psi_j u \in H_{\sigma+(j/2)}$ and $(\partial/\partial x_i)(\psi_j u) \in H_{\sigma+(j-1)/2}$ provided $\sigma + (j/2) \le s$.]

8. FOURIER ANALYSIS OF MEASURES

In this section all functions will be assumed to be Borel measurable. We recall that $M(\mathbf{R}^n)$ is the space of complex Borel measures on $\mathbf{R}^n$ [which are automatically Radon measures by Theorem (7.8)], and we embed $L^1(\mathbf{R}^n)$ into $M(\mathbf{R}^n)$ by identifying $f \in L^1$ with the measure $d\mu = f\,dm$. We shall need to define products of complex measures, which can easily be done in terms of products of positive measures by using Radon–Nikodym derivatives. Namely, if $\mu, \nu \in M(\mathbf{R}^n)$ we define $\mu \times \nu \in M(\mathbf{R}^n \times \mathbf{R}^n)$ by

$$d(\mu \times \nu)(x, y) = \frac{d\mu}{d|\mu|}(x)\frac{d\nu}{d|\nu|}(y)\,d(|\mu| \times |\nu|)(x, y).$$

Previously we have considered convolutions of functions with functions and with distributions. If $\mu, \nu \in M(\mathbf{R}^n)$ we now define their convolution $\mu * \nu \in M(\mathbf{R}^n)$ by $\mu * \nu(E) = \mu \times \nu(A^{-1}(E))$ where $A\colon \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ is the addition map $A(x, y) = x + y$. In other words,

$$\mu * \nu(E) = \iint \chi_E(x + y)\, d\mu(x)\, d\nu(y). \tag{8.66}$$

(8.67) Proposition.

(a) Convolution of measures is commutative and associative.

(b) For any bounded Borel measurable function h, $\int h\, d(\mu * \nu) = \iint h(x + y)\, d\mu(x)\, d\nu(y)$.

(c) $\|\mu * \nu\| \le \|\mu\| \|\nu\|$.

(d) If $d\mu = f\,dm$ and $d\nu = g\,dm$, then $d(\mu * \nu) = (f * g)\,dm$; that is, on L^1 the new and old definitions of convolution coincide.

Proof. Commutativity is obvious from Fubini's theorem, as is associativity, for $\lambda * \mu * \nu$ is unambiguously defined by the formula

$$\lambda * \mu * \nu(E) = \iiint \chi_E(x + y + z)\, d\lambda(x)\, d\mu(y)\, d\nu(z).$$

(b) follows from (8.66) by the usual linearity and approximation arguments. In particular, taking $h = d|\mu * \nu|/d(\mu * \nu)$, since $|h| = 1$ we obtain

$$\|\mu * \nu\| = \int h\, d(\mu * \nu) \le \iint d|\mu|(x)\, d|\nu|(y) = \|\mu\| \|\nu\|.$$

Finally, if $d\mu = f\,dm$ and $d\nu = g\,dm$, for any bounded h we have

$$\int h\, d(\mu * \nu) = \iint h(x + y) f(x) g(y)\, dx\, dy$$

$$= \iint h(x) f(x - y) g(y)\, dx\, dy = \int h(x)(f * g)(x)\, dx,$$

whence $d(\mu * \nu) = (f * g)\, dm$. □

We can also define convolutions of measures with L^p functions:

(8.68) Proposition. If $f \in L^p(\mathbf{R}^n)$ $(1 \le p \le \infty)$ and $\mu \in M(\mathbf{R}^n)$, the integral $f * \mu(x) = \int f(x - y)\, d\mu(y)$ exists for a.e. x, $f * \mu \in L^p$, and $\|f * \mu\|_p \le \|f\|_p \|\mu\|$. (Here "$L^p$" and "a.e." refer to Lebesgue measure.)

Proof. If f and μ are nonnegative, $f * \mu(x)$ exists for every x, possibly being equal to ∞, and by Minkowski's inequality for integrals,

$$\|f * \mu\|_p \le \int \|f(\cdot - y)\|_p\, d\mu(y) = \|f\|_p \|\mu\|.$$

In particular, $f * \mu(x) < \infty$ for a.e. x. In the general case this argument applies to $|f|$ and $|\mu|$, and the result follows easily. □

In the case $p = 1$ the definition of $f * \mu$ given in Proposition (8.68) coincides with the definition given earlier in which f is identified with $f\,dm$, for

$$\int_E f * \mu(x)\, dx = \iint \chi_E(x) f(x - y)\, d\mu(y)\, dx$$

$$= \iint \chi_E(x + y) f(x)\, dx\, d\mu(y).$$

Thus L^1 is not merely a subalgebra of $M(\mathbf{R}^n)$ with respect to convolution but an ideal. If $\mu \in M(\mathbf{R}^n)$, then μ is a tempered distribution and as such has a Fourier transform $\hat{\mu} \in \mathcal{S}'$. But we can also define $\hat{\mu}$ as a function by the formula

$$\hat{\mu}(E) = \int e^{-2\pi i x \cdot \xi}\, d\mu(x).$$

Since $e^{-2\pi i x \cdot \xi}$ is uniformly continuous in x, it is clear that $\hat{\mu}$, thus defined, is a bounded continuous function, and $\|\hat{\mu}\|_u \le \|\mu\|$. ($\hat{\mu}$ is often called the **Fourier–Stieltjes transform** of μ.) The discussion in Section 6 shows that the distribution and function definitions of $\hat{\mu}$ coincide when μ has compact support, and an easy limiting argument yields the validity of this result in general. Moreover, by taking $h(x) = e^{-2\pi i x \cdot \xi}$ in Proposition (8.67b) one sees immediately that $(\mu * \nu)\hat{} = \hat{\mu}\hat{\nu}$.

We conclude by giving a useful criterion for vague convergence of measures in terms of Fourier transforms.

(8.69) Proposition. If $\mu_1, \mu_2, \ldots,$ and μ are in $M(\mathbf{R}^n)$, $\sup_k \|\mu\|_k < \infty$, and $\hat{\mu}_k \to \hat{\mu}$ pointwise, then $\mu_k \to \mu$ vaguely.

Proof. If $f \in \mathcal{S}$, we have

$$\int f\, d\mu_k = \iint \hat{f}(y) e^{-2\pi i y \cdot x}\, dy\, d\mu_k(x) = \int \hat{f} \hat{\mu}_k.$$

Since $\check{f} \in \mathcal{S}$ and $\|\hat{\mu}_k\|_u \le \|\mu_k\|$, the dominated convergence theorem implies that $\int f\, d\mu_k \to \int f\, d\mu$. But $\mathcal{S}$ is dense in $C_0(\mathbf{R}^n)$, so by Proposition (5.17), $\int f\, d\mu_k \to \int f\, d\mu$ for all $f \in C_0(\mathbf{R}^n)$, that is, $\mu_k \to \mu$ vaguely. □

This result has a partial converse: if $\mu_k \to \mu$ vaguely and $\|\mu_k\| \to \|\mu\|$, then $\hat{\mu}_k \to \hat{\mu}$ pointwise. This follows from Exercise 7.26.

EXERCISES

56. Work out the analogs of the results in this section for measures on the torus $\mathbf{T}^n$.

57. If μ is a positive Borel measure on $\mathbf{T}^1$ with $\mu(\mathbf{T}^1) = 1$, then $|\hat{\mu}(k)| < 1$ for all $k \ne 0$ unless μ is a convex combination of the point masses at $0, m^{-1}, \ldots, (m - 1)m^{-1}$ for some $m \in \mathbf{N}$, in which case $\hat{\mu}(km) = 1$ for all $k \in \mathbf{Z}$.

58. $L^1(\mathbf{R}^n)$ is vaguely dense in $M(\mathbf{R}^n)$. [If $\mu \in M(\mathbf{R}^n)$, consider $\varphi_t * \mu$ where $\{\varphi_t\}$ is an approximate identity.]

59. If $\Delta(\mathbf{R}^n)$ is the set of finite linear combinations of the point masses δ_x, $x \in \mathbf{R}^n$, then $\Delta(\mathbf{R}^n)$ is vaguely dense in $M(\mathbf{R}^n)$. [If f is in the dense subset $C_c(\mathbf{R}^n)$ of $L^1(\mathbf{R}^n)$ and $g \in C_0(\mathbf{R}^n)$, approximate $\int fg$ by Riemann sums. Then use Exercise 58.]

60. If $1 \le p \le \infty$, let $\mathscr{C}_p$ be the space of all $F \in \mathscr{S}'$ such that $\|F * \varphi\|_p \le C\|\varphi\|_p$ for all $\varphi \in \mathscr{S}$ (with C independent of φ), so that the map $\varphi \to F * \varphi$ extends to a bounded operator on L^p.

(a) $\mathscr{C}_1 = M(\mathbf{R}^n)$. (If $F \in \mathscr{C}_1$, consider $F * \varphi_t$ where $\{\varphi_t\}$ is an approximate identity, and apply Alaoglu's theorem.)

(b) $\mathscr{C}_2 = \{F \in \mathscr{S}' : \hat{F} \in L^\infty\}$. (Use the Plancherel theorem.)

(c) If p and q are conjugate exponents with $1 < p < 2 < q < \infty$, then $\mathscr{C}_p = \mathscr{C}_q$. [Hint: $\langle F * \varphi, \psi \rangle = \langle F * \tilde{\psi}, \tilde{\varphi} \rangle$ where $\tilde{\varphi}(x) = \varphi(-x)$.]

(d) If p, q are as in (c), then $\mathscr{C}_p \subset \mathscr{C}_r$ for all $r \in [p, q]$. (In particular, $\mathscr{C}_1 \subset \mathscr{C}_p \subset \mathscr{C}_2$ for all p.)

9. NOTES AND REFERENCES

The scope of Fourier analysis is much wider than we have been able to indicate in this chapter. For a broader view of the subject, the reader is urged to consult the excellent collection of expository articles edited by Ash [4]. Also recommended is Dym and McKean [27], which discusses many interesting applications. On the more advanced level, the reader should consult Zygmund [113] for the classical theory, and Stein [100] and Stein and Weiss [101] for some of the more recently developed aspects of the subject.

Section 3. There is no universal agreement on where to put the factors of 2π in the definition of the Fourier transform. Other common conventions are

$$\mathscr{F}_1 f(\xi) = \int e^{-ix\cdot\xi} f(x)\, dx, \qquad \mathscr{F}_2 f(\xi) = (2\pi)^{-n/2} \int e^{-ix\cdot\xi} f(x)\, dx,$$

whose inverse transforms are, respectively,

$$\mathscr{F}_1^{-1} g(x) = (2\pi)^{-n} \int e^{ix\cdot\xi} g(\xi)\, d\xi, \qquad \mathscr{F}_2^{-1} g(x) = (2\pi)^{-n/2} \int e^{ix\cdot\xi} g(\xi)\, d\xi.$$

$\mathscr{F}_1$ has the disadvantage of not being unitary ($\|\mathscr{F}_1 f\|_2 = (2\pi)^{n/2}\|f\|_2$), while $\mathscr{F}_2$ is unitary but does not convert convolution into multiplication [$\mathscr{F}_2(f * g) = (2\pi)^{n/2}(\mathscr{F}_2 f)(\mathscr{F}_2 g)$]. One way to sweep all factors of 2π under the rug is to use the exponential $e^{-ix\cdot\xi}$ in defining $\mathscr{F}$ and replace Lebesgue measure m by $(2\pi)^{-n/2}m$ in defining all integrals (both Fourier transforms and convolutions)—but then one must remember to put the 2π's back in when computing specific integrals.

A nice complex-variable proof of the fact that the Fourier transform is injective on L^1 can be found in Newman [74].

Section 4. The study of pointwise convergence of one-dimensional Fourier series has a long and complex history; see Zygmund [113] and the articles by

Zygmund and Hunt in Ash [4]. Here are a few highlights:

(i) (Dirichlet, 1829) If f is piecewise continuous and piecewise monotone on $\mathbf{T}^1$, then $S_m f(x) \to \frac{1}{2}[f(x^+) + f(x^-)]$ for all x.

(ii) (du Bois Reymond, 1876) There exists $f \in C(\mathbf{T}^1)$ such that $\{S_m f(0)\}$ diverges.

(iii) (Fejér, 1904) If $f \in C(\mathbf{T}^1)$, then the Cesàro means of $\{S_m f\}$ converge uniformly to f.

(iv) (Kolmogorov, 1926) There exists $f \in L^1(\mathbf{T}^1)$ such that $\{S_m f(x)\}$ diverges for *every* x.

(v) (Carleson, 1966, and Hunt, 1967) If $f \in L^p(\mathbf{T}^1)$ where $p > 1$, then $S_m f(x) \to f(x)$ for almost every x.

An interesting historical discussion of the Gibb phenomenon can be found in Hewitt and Hewitt [52].

Section 5. The theory of distributions is the creation of Laurent Schwartz, although many of the ideas were implicit in earlier works; see Dieudonné [23]. More complete accounts of the subject can be found in Schwartz [96] and Treves [108]. In these books distributions are defined as the continuous linear functionals on C_c^∞ with respect to a certain topology. This topology on C_c^∞ is rather complicated (and not metrizable); for most purposes the characterization of distributions which we have adopted as a definition seems easier to work with.

Proposition (8.51) can be sharpened: if $F \in \mathscr{E}'$, then $\hat{F}(\xi) = \langle F, E_\xi \rangle$ is well defined for all $\xi \in \mathbf{C}^n$ and is an entire holomorphic function on $\mathbf{C}^n$.

Section 6. One can define L^p Sobolev spaces H_s^p for any $p \in [1, \infty]$. If $s \in \mathbf{N}$, H_s^p is the set of all $f \in L^p$ such that $\partial^\alpha f \in L^p$ for $|\alpha| \le s$. When $1 < p < \infty$, it turns out that $H_s^p = \{f : \Lambda_s f \in L^p\}$ (although this is not trivial to prove when $p \ne 2$), and this characterization of H_s^p can be used as a definition for all $s \in \mathbf{R}$. See Stein [100], Chapter V.

In Rellich's theorem the hypothesis that V is compact can be replaced by the hypothesis that $m(V) < \infty$, provided $s \ge 0$. See Lair [60].

Section 7. A differential operator $L = \sum_{|\alpha| \le m} a_\alpha D^\alpha$ with C^∞ coefficients is called **elliptic** on Ω if $\sum_{|\alpha|=m} a_\alpha(x)\xi^\alpha \ne 0$ for all $\xi \ne 0$ and $x \in \Omega$. The elliptic regularity theorem remains true as stated for such operators; see Folland [32]. If L is elliptic on Ω, $u \in \mathscr{D}'$, and $Lu \in C^k(\Omega)$, it is in general false that $u \in C^{k+m}(\Omega)$ except in dimension 1. The correct statement is that if $Lu \in C^k(\Omega)$ and $\partial^\alpha Lu\,(|\alpha| = k)$ is locally Hölder continuous on Ω of exponent a, $0 < a < 1$, then $u \in C^{k+m}(\Omega)$ and $\partial^\alpha u$ $(|\alpha| = k + m)$ satisfies the same Hölder condition. Moreover, the L^p version of the elliptic regularity theorem is true for $1 < p < \infty$. See Taylor [105].

Further applications of Fourier analysis to partial differential equations can be found, for example, in Folland [32, 33], and Taylor [105].

9

TOPICS IN PROBABILITY THEORY

The fundamental datum in probability theory is a measure space $(X, \mathscr{M}, \mu)$ such that $\mu(X) = 1$; such a measure μ is called a **probability measure**. X is to be considered as the set of all possible outcomes of some process, such as an experiment or a gambling game; and the measure of a set $E \in \mathscr{M}$ is interpreted as the probability that the outcome lies in E. Although measure spaces are the natural setting for the study of probability, it is hardly accurate to say that probability theory is a branch of measure theory, for its central ideas and many of its techniques are distinctively its own. This brief chapter is intended not as a systematic introduction to probability theory but rather as an advertisement for the subject; it also serves to illustrate further some results of previous chapters.

1. BASIC CONCEPTS

Probability theory has its own vocabulary, which is partly a legacy of its development before the connection with measure theory was made explicit and partly a result of the fact that the probabilistic point of view is different. We therefore begin by presenting a brief dictionary of probabilists' dialect.

Analysts' Term	*Probabilists' Term*
Measure space with total measure 1. Notation: $(X, \mathscr{M}, \mu)$	Sample space or probability space. Notation: $(\Omega, \mathscr{B}, P)$
(σ-)algebra	(σ-)field
Measurable set	Event
Measurable real-valued function. Notation: f	Random variable. Notation: X

Integral of f. Notation: $\int f d\mu$	Expectation, or expected value, or mean, of X. Notation: $E(X)$
L^p function	Random variable having finite pth moment
Convergence in measure	Convergence in probability
Almost every(where) (a.e.)	Almost sure(ly) (a.s.)
Borel probability measure on $\mathbf{R}$	Distribution
Fourier transform (of a measure on $\mathbf{R}$)	Characteristic function (of a distribution)
Characteristic function	Indicator function

Probabilists have an aversion to displaying the arguments of random variables. For example, $\{\omega\colon X(\omega) > \alpha\}$ and $P(\{\omega\colon X(\omega) > \alpha\})$ are commonly written as $\{X > \alpha\}$ and $P(X > \alpha)$.

Henceforth we shall, for the most part, adopt probabilistic language in this chapter, although we shall use the term "L^p random variable" in preference to the more cumbersome "random variable with finite pth moment." One more standard piece of terminology, which has no equivalent in classical analysis, is the following. If X is an L^2 random variable, then $X \in L^1$ since P is a finite measure, so that the mean $E(X)$ is well defined. The L^2 norm of $X - E(X)$ is called the **standard deviation** of X, and its square is called the **variance** of X. These quantities are denoted by $\sigma(X)$ and $\sigma^2(X)$. Thus:

$$\sigma^2(X) = E\left[(X - E(X))^2\right] = E(X^2) - E(X)^2, \qquad \sigma(X) = \sqrt{\sigma^2(X)}\,.$$

$\sigma(X)$ is a measure of how widely X deviates from its mean value.

At this point we must discuss a general measure-theoretic construction. Let $(\Omega, \mathscr{B}, P)$ be a probability space (or, for that matter, an arbitrary measure space), let $(\Omega', \mathscr{B}')$ be a measurable space, and let $\varphi\colon \Omega \to \Omega'$ be a $(\mathscr{B}, \mathscr{B}')$-measurable map. Then the measure P induces an **image measure** P_φ on Ω' by

$$P_\varphi(E) = P\left(\varphi^{-1}(E)\right).$$

That this is indeed a measure follows from the fact that φ^{-1} commutes with unions and intersections.

(9.1) Proposition. With notation as above, if $f\colon \Omega' \to \mathbf{R}$ is a measurable function then $\int_{\Omega'} f\, dP_\varphi = \int_\Omega (f \circ \varphi)\, dP$ whenever either side is defined.

Proof. When $f = \chi_E$ for $E \in \mathscr{B}'$ this is just the definition of P_φ, since $\chi_E \circ \varphi = \chi_{\varphi^{-1}(E)}$. The general result follows by taking linear combinations and limits. □

If X is a random variable on Ω, then P_X is a Borel probability measure on $\mathbf{R}$, called the **distribution** of X, and the function

$$F(t) = P_X((-\infty, t]) = P(X \le t)$$

[which determines P_X by Theorem (1.16)] is called the **distribution function** of X. More generally, if $X_1, \dots, X_n$ are random variables, we can consider $(X_1, \dots, X_n)$ as a map from Ω to $\mathbf{R}^n$, and the measure $P_{(X_1,\dots,X_n)}$ on $\mathbf{R}^n$ is called the **joint distribution** of $X_1, \dots, X_n$. It is a general principle that all properties of random variables which are relevant to probability theory can be expressed in terms of their joint distributions. For example, by Proposition 9.1,

$$E(X) = \int t\, dP_X(t), \quad \sigma^2(X) = \int (t - E(X))^2\, dP_X(t),$$

$$E(X + Y) = \int (t + s)\, dP_{(X,Y)}(t, s).$$

If $\{X_\alpha\}$ is a family of random variables such that $P_{X_\alpha} = P_{X_\beta}$ for all α, β, the X_α's are said to be **identically distributed**. In this case the mean and variance of X_α are independent of α.

One of the most important concepts in probability theory, and the one which most clearly sets it apart from general measure theory, is that of (stochastic) independence. To motivate this idea, consider a probability space $(\Omega, \mathscr{B}, P)$ and an event E such that $P(E) > 0$. Then the set function $P_E(F) = P(E \cap F)/P(E)$ is a probability measure on Ω called the conditional probability on E; $P_E(F)$ represents the probability of the event F given that E occurs. If $P_E(F) = P(F)$, that is, if the probability of F is the same whether or not we restrict to E, then F is said to be independent of E. This condition is equivalent to the condition $P(E \cap F) = P(E)P(F)$, which is clearly symmetric in E and F and makes sense even if $P(E) = 0$.

With this in mind, we define a collection $\{E_\alpha\}_{\alpha \in A}$ of events in Ω to be **independent** if

$$P(E_{\alpha_1} \cap \cdots \cap E_{\alpha_n}) = \prod_1^n P(E_{\alpha_j}) \quad \text{for all } \alpha_1, \dots, \alpha_n \in A, n \in \mathbf{N}.$$

(For the events E_α to be independent it does *not* suffice for them to be pairwise independent: see Exercise 1.) A collection $\{X_\alpha\}_{\alpha \in A}$ of random variables on Ω is called **independent** if the events $\{X_\alpha \in B_\alpha\} = X_\alpha^{-1}(B_\alpha)$ are independent for all Borel sets $B_\alpha \subset \mathbf{R}$. This condition can be neatly rephrased as follows. Observe that if $\alpha_1, \dots, \alpha_n \in A$ and we write $X_j = X_{\alpha_j}$, we have

$$P(X_1^{-1}(B_1) \cap \cdots \cap X_n^{-1}(B_n)) = P((X_1, \dots, X_n)^{-1}(B_1 \times \cdots \times B_n))$$
$$= P_{(X_1,\dots,X_n)}(B_1 \times \cdots \times B_n),$$

whereas

$$\prod_1^n P(X_j^{-1}(B_j)) = \prod_1^n P_{X_j}(B_j) = \left(\prod_1^n P_{X_j}\right)(B_1 \times \cdots \times B_n).$$

Since the B_j's are arbitrary Borel sets in $\mathbf{R}$, these quantities are equal iff

$$P_{(X_1,\dots,X_n)} = \prod_1^n P_{X_j}.$$

That is, *the random variables X_α are independent iff their joint distributions are the products of their individual distributions.*

The following proposition expresses the fact that functions of independent random variables are independent.

(9.2) Proposition. Let $\{X_{nj}: 1 \le j \le J(n), 1 \le n \le N\}$ be independent random variables, and let $f_n: \mathbf{R}^{J(n)} \to \mathbf{R}$ be Borel measurable for $1 \le n \le N$. Then the random variables $Y_n = f_n(X_{n1}, \dots, X_{nJ(n)})$ $(1 \le n \le N)$ are independent.

Proof. Let $\boldsymbol{X}_n = (X_{n1}, \dots, X_{nJ(n)})$. Then if $B_1, \dots, B_N$ are Borel subsets of $\mathbf{R}$ we have $Y_n^{-1}(B_n) = \boldsymbol{X}_n^{-1}(f_n^{-1}(B_n))$ and hence

$$(Y_1, \dots, Y_N)^{-1}(B_1 \times \cdots \times B_N)$$
$$= \bigcap_1^N Y_n^{-1}(B_n) = (\boldsymbol{X}_1, \dots, \boldsymbol{X}_N)^{-1}(f_1^{-1}(B_1) \times \cdots \times f_N^{-1}(B_N)).$$

Therefore, by the independence of the X_{nj}'s and Fubini's theorem,

$$P_{(Y_1,\dots,Y_N)}(B_1 \times \cdots \times B_N) = P_{(\boldsymbol{X}_1,\dots,\boldsymbol{X}_N)}(f_1^{-1}(B_1) \times \cdots \times f_N^{-1}(B_N))$$
$$= \left(\prod_{n=1}^N \prod_{j=1}^{J(n)} P_{X_{nj}}\right)(f_1^{-1}(B_1) \times \cdots \times f_N^{-1}(B_N))$$
$$= \prod_{n=1}^N P_{\boldsymbol{X}_n}(f_n^{-1}(B_n))$$
$$= \prod_{n=1}^N P_{Y_n}(B_n). \qquad \square$$

We now present some fundamental properties of independent random variables. For the first one we need the notion of convolutions of measures on $\mathbf{R}$ developed in Section 8.8. An easy induction on (8.66) shows that if $\lambda_1, \dots, \lambda_n \in M(\mathbf{R})$, then $\lambda_1 * \cdots * \lambda_n$ is given by

$$(9.3) \quad \lambda_1 * \cdots * \lambda_n(E) = \int \cdots \int \chi_E(t_1 + \cdots + t_n)\, d\lambda_1(t_1) \cdots d\lambda_n(t_n).$$

(9.4) Proposition. If $\{X_j\}_1^n$ are independent random variables, then

$$P_{X_1+\cdots+X_n} = P_{X_1} * \cdots * P_{X_n}.$$

Proof. Let $A(t_1,\ldots,t_n) = \sum_1^n t_j$. Then $X_1 + \cdots + X_n = A \circ (X_1,\ldots,X_n)$, so

$$P_{X_1+\cdots+X_n} = \left(P_{(X_1,\ldots,X_n)}\right)_A = \left(\prod_1^n P_{X_j}\right)_A,$$

and, by (9.3), the last quantity equals $P_{X_1} * \cdots * P_{X_n}$. □

(9.5) Proposition. Suppose $\{X_j\}_1^n$ are independent random variables. If $X_j \in L^1$ for all j, then $\prod_1^n X_j \in L^1$, and $E(\prod_1^n X_j) = \prod_1^n E(X_j)$.

Proof. We have $\prod_1^n |X_j| = f(X_1,\ldots,X_n)$ where $f(t_1,\ldots,t_n) = \prod_1^n |t_j|$. Hence

$$E\left(\prod_1^n |X_j|\right) = \int f\,dP_{(X_1,\ldots,X_n)} = \int f\,d\left(\prod_1^n P_{X_j}\right)$$

$$= \prod_1^n \int |t_j|\,dP_{X_j}(t_j) = \prod_1^n E(|X_j|).$$

This proves the first assertion, and once this is known the same argument, with the absolute values removed, proves the second one. □

(9.6) Corollary. If $\{X_j\}_1^n$ are independent and in L^2, then $\sigma^2(X_1 + \cdots + X_n) = \sum_1^n \sigma^2(X_j)$.

Proof. Let $Y_j = X_j - E(X_j)$. Then $\{Y_j\}_1^n$ are independent and have mean zero, so

$$E(Y_jY_k) = E(Y_j)E(Y_k) = 0 \quad (j \neq k).$$

Therefore,

$$\sigma^2(X_1 + \cdots + X_n) = E\left((Y_1 + \cdots + Y_n)^2\right) = \sum_{j,k} E(Y_jY_k)$$

$$= \sum_j E(Y_j^2) = \sum_j \sigma^2(X_j).$$ □

These results show that independence is a very stringent property. For one thing, it is usually not the case that the product of two integrable functions is

integrable. For another, suppose X and Y are independent and $E(X) = 0$. Then for any Borel measurable function f on $\mathbf{R}$ such that $f \circ Y \in L^1$ we have

$$E(X(f \circ Y)) = E(X)E(f \circ Y) = 0.$$

In other words, X is orthogonal (in the L^2 sense) to *every* function of Y. This indicates that if, for example, one tries to construct a sequence of independent random variables on $[0,1]$ with Lebesgue measure by using the familiar functions of calculus, one will probably not succeed. (Try it!—then see Exercises 23–25, where an independent sequence on $[0,1]$ is displayed.) Rather, the natural setting for independence is product spaces. Suppose

$$\Omega = \Omega_1 \times \cdots \times \Omega_n, \quad \mathscr{B} = \mathscr{B}_1 \otimes \cdots \otimes \mathscr{B}_n, \quad P = P_1 \times \cdots \times P_n.$$

Then any random variables $X_1,\ldots,X_n$ on Ω such that X_j depends only on the jth coordinate are independent, for if $X_j = f_j \circ \pi_j$ where $\pi_j\colon \Omega \to \Omega_j$ is the projection,

$$(X_1,\ldots,X_n)^{-1}(B_1 \times \cdots \times B_n) = \prod_1^n f_j^{-1}(B_j)$$

and hence

$$P_{(X_1,\ldots,X_n)}(B_1 \times \cdots \times B_n) = P\left(\prod_1^n f_j^{-1}(B_j)\right)$$

$$= \prod_1^n P_j\left(f_j^{-1}(B_j)\right) = \prod_1^n P_{X_j}(B_j).$$

The same idea can be made to work for infinitely many factors: see Section 4.

As an application of these ideas we present Bernstein's constructive proof of the Weierstrass approximation theorem. If $f \in C([0,1])$, the nth **Bernstein polynomial** of f is

$$B_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right)\frac{n!}{k!(n-k)!}x^k(1-x)^{n-k}.$$

(9.7) Theorem. With notation as above, $B_n \to f$ uniformly on $[0,1]$ as $n \to \infty$.

Proof. Given $x \in [0,1]$, let $\lambda = x\delta_1 + (1-x)\delta_0$ where δ_t is the point mass at t. Let $\Omega = \mathbf{R}^n$, $P = \lambda \times \cdots \times \lambda$, and $X_j =$ the jth coordinate function on $\mathbf{R}^n$. Then $X_1,\ldots,X_n$ are independent and have the common distribution λ. It is easy to check (Exercise 7) that the distribution of $X_1 + \cdots + X_n$ is

$$\sum_0^n \frac{n!}{k!(n-k)!}x^k(1-x)^{n-k}\delta_k,$$

and hence

$$B_n(x) = E\left[f\left(\frac{X_1 + \cdots + X_n}{n}\right)\right].$$

Now, $\Sigma_0^n n![k!(n-k)!]^{-1}x^k(1-x)^{n-k} = 1$ by the binomial theorem, so

$$|f(x) - B_n(x)| \le \sum_0^n \left|f(x) - f\left(\frac{k}{n}\right)\right| \frac{n!}{k!(n-k)!}x^k(1-x)^{n-k}. \tag{9.8}$$

Given $\varepsilon > 0$, choose $\delta > 0$ such that $|f(x) - f(y)| \le \varepsilon$ whenever $|x - y| \le \delta$. Then the sum of the terms in (9.8) such that $|x - (k/n)| \le \delta$ is at most ε, while the sum of the remaining terms is at most

$$2\|f\|_u P\left(\left|\frac{X_1 + \cdots + X_n}{n} - x\right| > \delta\right).$$

But

$$\sigma^2(X_j) = E(X_j^2) - E(X_j)^2 = x - x^2 \le 1,$$

so by Corollary (9.6),

$$E\left[\left(\frac{X_1 + \cdots + X_n}{n} - x\right)^2\right] = \sigma^2\left(\frac{X_1 + \cdots + X_n}{n}\right) \le \frac{n}{n^2} = \frac{1}{n},$$

and Chebyshev's inequality therefore gives

$$|f(x) - B_n(x)| \le \varepsilon + \frac{2\|f\|_u}{n\delta^2},$$

which is less than 2ε provided n is sufficiently large. □

EXERCISES

1. Let Ω consist of four points, each with probability $\frac{1}{4}$. Find three events which are pairwise independent but not independent. Generalize.
2. Let $\{X_j\}$ be a sequence of independent identically distributed positive random variables such that $E(X_j) = a < \infty$ and $E(1/X_j) = b < \infty$, and let $S_n = \Sigma_1^n X_j$.
   (a) $E(X_j/S_n)$ equals n^{-1} if $j \le n$ and $aE(1/S_n)$ if $j > n$.
   (b) $E(S_m/S_n)$ equals mn^{-1} if $m \le n$ and $1 + (m-n)aE(1/S_n)$ if $m > n$.

3. (a) If $E_1, \ldots, E_n$ are independent events, so are $E_1, \ldots, E_{n-1}, E_n^c$.
   (b) If $\{E_\alpha\}$ is an independent set of events, so is $\{F_\alpha\}$ where each F_α is either E_α or E_α^c. [Use induction on (a).]
   (c) The events $\{E_\alpha\}$ are independent iff the random variables $\{\chi_{E_\alpha}\}$ are independent.
4. Let X, Y, Z be positive independent random variables with a common distribution λ, and let $F(t) = \lambda((0,t])$. The probability that the polynomial $Xt^2 + Yt + Z$ has real roots is $\int_0^\infty \int_0^\infty F(t^2/4s)\, d\lambda(t)\, d\lambda(s)$.
5. If X is a random variable such that $dP_X(t) = f(t)\, dt$ where $f(t) = f(-t)$, then $dP_{X^2}(t) = t^{-1/2}f(t^{1/2})\chi_{(0,\infty)}(t)\, dt$.

The next three exercises deal with distributions, that is, probability measures on $\mathbf{R}$. If λ is a distribution, its **mean** is $\mu = \int t\, d\lambda(t)$ and its **variance** is $\sigma^2 = \int (t-\mu)^2\, d\lambda(t) = \int t^2\, d\lambda(t) - \mu^2$. If $d\lambda(t) = f(t)\, dt$, f is called the **density** of λ. δ_t denotes the point mass at t.

6. If $a, u > 0$, the distribution $\gamma_{a,u}$ whose density is $[\Gamma(a)]^{-1}u^a t^{a-1}e^{-ut}\chi_{(0,\infty)}(t)$ is called a **gamma distribution**.
   (a) The mean and variance of $\gamma_{a,u}$ are a/u and a/u^2, respectively.
   (b) $\gamma_{a,u} * \gamma_{b,u} = \gamma_{a+b,u}$. (Use Exercise 2.62.)
   (c) If $X_1, \ldots, X_n$ are independent and distributed with density $(2\pi)^{-1/2}e^{-t^2/2}$, then $X_1^2 + \cdots + X_n^2$ has the distribution $\gamma_{n/2,1/2}$. [Use (b) and Exercise 5. $\gamma_{n/2,1/2}$ is called the **chi-square distribution** with n degrees of freedom.]
7. If $0 < p < 1$, let $\beta_p = p\delta_1 + (1-p)\delta_0$, and let $\beta_p^{(n)}$ be the nth convolution power of β_p. Then

$$\beta_p^{(n)} = \sum_{k=0}^n \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}\delta_k.$$

   Also, the mean and variance of $\beta_p^{(n)}$ are np and $np(1-p)$. $\beta_p^{(n)}$ is called the **binomial distribution** on $\{0, \ldots, n\}$ with parameter p.
8. If $a > 0$, let

$$\lambda_a = e^{-a}\sum_0^\infty \frac{a^k}{k!}\delta_k.$$

   λ_k is called the **Poisson distribution** with parameter a.
   (a) The mean and variance of λ_a are both a.
   (b) $\lambda_a * \lambda_b = \lambda_{a+b}$.
   (c) $\beta_{a/n}^{(n)}$ (as in Exercise 7) converges vaguely to λ_a as $n \to \infty$. [Use Proposition (7.19).]

2. THE LAW OF LARGE NUMBERS

If one plays a gambling game many times, one's average winnings or losses per game should be roughly the expected winnings or losses in each individual game; more generally, if one plays a sequence of possibly different games, one's average winnings or losses should be roughly the average of the expected winnings or losses in the individual games. The law of large numbers is a precise formulation of this idea. It comes in several versions, depending on the hypothesis one wishes to make. The first version, with the weakest hypotheses and conclusions, has a very simple proof.

(9.9) The Weak Law of Large Numbers. Let $\{X_j\}_1^\infty$ be a sequence of independent L^2 random variables with means $\{\mu_j\}$ and variances $\{\sigma_j^2\}$. If $\lim n^{-2}\Sigma_1^n\sigma_j^2 = 0$, then $n^{-1}\Sigma_1^n(X_j - \mu_j) \to 0$ in probability as $n \to \infty$.

Proof. $n^{-1}\Sigma_1^n(X_j - \mu_j)$ has mean zero and variance $n^{-2}\Sigma_1^n\sigma_j^2$ [the latter by Corollary (9.6)]. Hence by Chebyshev's inequality, for any $\varepsilon > 0$ we have

$$P\left(\left|n^{-1}\sum_1^n (X_j - \mu_j)\right| > \varepsilon\right) \le (n\varepsilon)^{-2}\sum_1^n \sigma_j^2 \to 0 \quad \text{as } n \to \infty. \qquad \square$$

Under slightly stronger hypotheses, one obtains the sharper conclusion that $n^{-1}\Sigma_1^n(X_j - \mu_j) \to 0$ almost surely. To establish this, we need the following two lemmas, which are of interest in their own right.

(9.10) The Borel–Cantelli Lemma. Let $\{A_n\}$ be a sequence of events. If $\Sigma_1^\infty P(A_n) < \infty$, then $P(\limsup A_n) = 0$. If the A_n's are independent and $\Sigma_1^\infty P(A_n) = \infty$, then $P(\limsup A_n) = 1$.

Proof. We recall that $\limsup A_n = \bigcap_{k=1}^\infty \bigcup_{n=k}^\infty A_n$, so that

$$P(\limsup A_n) \le P\left(\bigcup_{n=k}^\infty A_n\right) \le \sum_{n=k}^\infty P(A_n),$$

and the latter sum tends to zero if $\Sigma_1^\infty P(A_n)$ converges. On the other hand, suppose $\Sigma_1^\infty P(A_n)$ diverges and the A_n's are independent. We must show that

$$P\big((\limsup A_n)^c\big) = P\left(\bigcup_{k=1}^\infty \bigcap_{n=k}^\infty A_n^c\right) = 0,$$

and for this it is enough to show that $P(\bigcap_{n=k}^\infty A_n^c) = 0$ for all k. But the A_n^c's are independent (Exercise 3), so since $1 - t \le e^{-t}$,

$$P\left(\bigcap_{n=k}^K A_n^c\right) = \prod_k^K (1 - P(A_n)) \le \prod_k^K e^{-P(A_n)} = \exp\left(-\sum_k^K P(A_n)\right).$$

The last expression tends to zero as $K \to \infty$, which yields the desired result. $\square$

(9.11) Kolmogorov's Inequality. Let $X_1, \dots, X_n$ be independent random variables with mean 0 and variances $\sigma_1^2, \dots, \sigma_n^2$, and let $S_k = X_1 + \cdots + X_k$. For any $\varepsilon > 0$,

$$P(\max_{1 \le k \le n}|S_k| \ge \varepsilon) \le \varepsilon^{-2}\sum_1^n \sigma_k^2.$$

Proof. Let A_k be the set where $|S_j| < \varepsilon$ for $j < k$ and $|S_k| \ge \varepsilon$. Then the A_k's are disjoint and their union is the set where $\max|S_k| \ge \varepsilon$, so

$$P(\max|S_k| \ge \varepsilon) = \sum_1^n P(A_k) \le \varepsilon^{-2}\sum_1^n E\big(\chi_{A_k}S_k^2\big)$$

because $S_k^2 \ge \varepsilon^2$ on A_k. On the other hand,

$$\begin{aligned} E\big(S_n^2\big) &\ge \sum_1^n E\big(\chi_{A_k}S_n^2\big) \\ &= \sum_1^n E\big[\chi_{A_k}\big[S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2\big]\big] \\ &\ge \sum_1^n E\big(\chi_{A_k}S_k^2\big) + 2\sum_1^n E\big[\chi_{A_k}S_k(S_n - S_k)\big]. \end{aligned}$$

It will suffice to show that $E[\chi_{A_k}S_k(S_n - S_k)] = 0$ for all k, for then we have

$$P(\max|S_k| \ge \varepsilon) \le \varepsilon^{-2}E\big(S_n^2\big) = \varepsilon^{-2}\sum_1^n \sigma_k^2$$

by Corollary (9.6), since the X_k's have mean 0. But χ_{A_k} is a measurable function of $S_1, \dots, S_k$ and hence of $X_1, \dots, X_k$, whereas $S_n - S_k$ is a measurable function of $X_{k+1}, \dots, X_n$. Moreover, $E(S_k) = \Sigma_1^k E(X_j) = 0$ for all k. Therefore, by Propositions (9.2) and (9.5),

$$E\big[\chi_{A_k}S_k(S_n - S_k)\big] = E\big(\chi_{A_k}S_k\big)E(S_n - S_k) = E\big(\chi_{A_k}S_k\big)\cdot 0 = 0. \qquad \square$$

(9.12) Kolmogorov's Strong Law of Large Numbers. If $\{X_n\}$ is a sequence of independent L^2 random variables with means $\{\mu_n\}$ and variances $\{\sigma_n^2\}$ such that $\Sigma_1^\infty n^{-2}\sigma_n^2 < \infty$, then $n^{-1}\Sigma_1^n(X_j - \mu_j) \to 0$ almost surely as $n \to \infty$.

Proof. Let $S_n = \Sigma_1^n(X_j - \mu_j)$. Given $\varepsilon > 0$, for $k \in \mathbf{N}$ let A_k be the set where $n^{-1}|S_n| \ge \varepsilon$ for some n such that $2^{k-1} \le n < 2^k$. Then on A_k we have $|S_n| \ge \varepsilon 2^{k-1}$ for some $n < 2^k$, so by Kolmogorov's inequality,

$$P(A_k) \le (\varepsilon 2^{k-1})^{-2}\sum_1^{2^k} \sigma_n^2.$$

Therefore,

$$\sum_1^\infty P(A_k) \le \frac{4}{\varepsilon^2} \sum_{k=1}^\infty \sum_{n=1}^{2^k} 2^{-2k}\sigma_n^2$$

$$= \frac{4}{\varepsilon^2} \sum_{n=1}^\infty \Big(\sum_{k \ge \log_2 n} 2^{-2k} \Big) \sigma_n^2 \le \frac{8}{\varepsilon^2} \sum_{n=1}^\infty \frac{\sigma_n^2}{n^2} < \infty,$$

so $P(\limsup A_k) = 0$ by the Borel–Cantelli lemma. But $\limsup A_k$ is precisely the set where $n^{-1}|S_n| \ge \varepsilon$ for infinitely many n, so

$$P\big(\limsup n^{-1}|S_n| < \varepsilon\big) = 1.$$

Letting $\varepsilon \to 0$ through a countable sequence of values, we conclude that $n^{-1}S_n \to 0$ almost surely. □

The hypotheses of this theorem are a bit stronger than those of the weak law (Exercise 9). They are certainly satisfied when the X_n's are identically distributed independent L^2 random variables, since then σ_n^2 is independent of n. However, in the identically distributed case the assumption that $X_n \in L^2$ can be weakened.

(9.13) Khinchine's Strong Law of Large Numbers. If $\{X_n\}$ is a sequence of independent identically distributed L^1 random variables with mean μ, then $n^{-1}\sum_1^n X_j \to \mu$ almost surely as $n \to \infty$.

Proof. Replacing X_j by $X_j - \mu$, we may assume that $\mu = 0$. Let λ be the common distribution of the X_j's; we are thus assuming that

$$\int |t|\, d\lambda(t) < \infty, \qquad \int t\, d\lambda(t) = 0.$$

Also, let $Y_j = X_j$ on the set where $|X_j| \le j$ and $Y_j = 0$ elsewhere. Then

$$\sum_1^\infty P\big(Y_j \ne X_j\big) = \sum_1^\infty P\big(|X_j| > j\big) = \sum_1^\infty \lambda\big(\{t\colon |t| > j\}\big)$$

$$= \sum_1^\infty j\lambda\big(\{t\colon j < |t| \le j+1\}\big) \le \int |t|\, d\lambda(t) < \infty.$$

By the Borel–Cantelli lemma, then, with probability one we have $X_j = Y_j$ for j sufficiently large, and it therefore suffices to show that $n^{-1}\sum_1^n Y_j \to 0$ almost surely. We have

$$\sigma^2(Y_n) \le E\big(Y_n^2\big) = \int_{|t| \le n} t^2\, d\lambda(t),$$

and hence

$$\sum_1^\infty n^{-2}\sigma^2(Y_n) \le \sum_{n=1}^\infty \sum_{j=1}^n n^{-2} \int_{j-1<|t|\le j} t^2\, d\lambda(t)$$

$$\le \sum_{n=1}^\infty \sum_{j=1}^n jn^{-2} \int_{j-1<|t|\le j} |t|\, d\lambda(t)$$

$$= \sum_{j=1}^\infty \sum_{n=j}^\infty jn^{-2} \int_{j-1<|t|\le j} |t|\, d\lambda(t).$$

But $\sum_{n=j}^\infty n^{-2} \le 2j^{-1}$ by comparison to the integral $\int_j^\infty x^{-2}\, dx$, so

$$\sum_1^\infty n^{-2}\sigma^2(Y_n) \le 2\sum_{j=1}^\infty \int_{j-1<|t|\le j} |t|\, d\lambda(t) = 2\int_{-\infty}^\infty |t|\, d\lambda(t) < \infty.$$

By Theorem (9.12), therefore, if $\mu_j = E(Y_j)$ we have

$$n^{-1}\sum_1^n Y_j - n^{-1}\sum_1^n \mu_j \to 0 \text{ almost surely.}$$

However, by the dominated convergence theorem,

$$\mu_j = \int_{|t|\le j} t\, d\lambda(t) \to \int_{-\infty}^\infty t\, d\lambda(t) = 0,$$

and it follows easily (Exercise 10) that $n^{-1}\sum_1^n \mu_j \to 0$ also. The proof is therefore complete. □

Thus far we have not shown how to construct sequences of random variables satisfying the hypotheses of the theorems in this section. We shall do so in Section 4.

EXERCISES

9. If $\sum_1^\infty n^{-2}\sigma_n^2 < \infty$, then $\lim n^{-2}\sum_1^n \sigma_j^2 = 0$. (Estimate $\sum_{j \le \varepsilon n} \sigma_j^2$ and $\sum_{j > \varepsilon n} \sigma_j^2$ separately.)

10. If $\{a_n\} \subset \mathbf{C}$ and $\lim a_n = a$, then $\lim n^{-1}\sum_1^n a_j = a$.

11. The weak law of large numbers remains valid if the hypothesis of independence is replaced by the (much weaker) hypothesis that $E[(X_j - \mu_j)(X_k - \mu_k)] = 0$ for $j \ne k$.

12. If $\{X_n\}$ is a sequence of independent random variables such that $E(X_n) = 0$ and $\sum_1^\infty \sigma^2(X_n) < \infty$, then $\sum_1^\infty X_n$ converges almost surely. (Apply

Kolmogorov's inequality to show that the partial sums are Cauchy a.s.) Corollary: if the plus and minus signs in $\sum_1^\infty \pm n^{-1}$ are determined by successive tosses of a fair coin, the resulting series converges almost surely.

13. If $\{X_n\}$ is a sequence of independent identically distributed random variables which are not in L^1, then $\limsup n^{-1}|\sum_1^n X_j| = \infty$ almost surely. [Let $A_n = \{|X_n| > n\}$. Using an idea from the proof of Theorem (9.13), show that $\sum_1^\infty P(A_n) = \infty$, and apply the Borel–Cantelli lemma.]

14. (Shannon's Theorem) Let (X_i) be a sequence of independent random variables on the sample space Ω having the common distribution $\lambda = \sum_1^r p_j \delta_j$ where $0 < p_j < 1$, $\sum_1^r p_j = 1$, and δ_j is the point mass at j. Define random variables $Y_1, Y_2, \ldots$ on Ω by

$$Y_n(\omega) = P\Big(\big\{\omega' \colon X_j(\omega') = X_j(\omega) \text{ for } 1 \le j \le n\big\}\Big).$$

(a) $Y_n = \prod_1^n p_{X_j}$.

(b) $n^{-1}\log Y_n \to \sum_1^r p_j \log p_j$ almost surely. (In information theory, the X_j's are considered as the output of a source of signals, and $-\sum_1^r p_j \log p_j$ is called the entropy of the signal.)

15. A collection or "population" of N objects (such as mice, snowflakes, etc.) may be considered as a sample space in which each object has probability N^{-1}. Let X be a random variable on this space—that is, a numerical characteristic of the objects such as weight, diameter, etc.—with mean μ and variance σ. In statistics one is interested in determining μ and σ^2 empirically by taking a sequence of random samples from the population and measuring X for each sample, thus obtaining a sequence $\{X_j\}$ of numbers which are values of independent random variables with the same distribution as X. $M_n = n^{-1}\sum_1^n X_j$ is called the nth **sample mean** and $S_n^2 = (n-1)^{-1}\sum_1^n (X_j - M_n)^2$ is called the nth **sample variance**. Show that $E(M_n) = \mu$, $E(S_n^2) = \sigma^2$, and $M_n \to \mu$ and $S_n^2 \to \sigma^2$ almost surely as $n \to \infty$. Can you see why one uses $(n-1)^{-1}$ instead of n^{-1} in the definition of S_n^2?

3. THE CENTRAL LIMIT THEOREM

If $\mu \in \mathbf{R}$ and $\sigma > 0$ the measure $\nu_\mu^{\sigma^2}$ on $\mathbf{R}$ defined by

$$d\nu_\mu^{\sigma^2}(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{(t-\mu)^2/2\sigma}\, dt$$

is a probability measure which satisfies

$$\int t\, d\nu_\mu^{\sigma^2} = \mu, \quad \int (t-\mu)^2\, d\nu_\mu^{\sigma^2}(t) = \sigma^2.$$

It is called the **normal** or **Gaussian distribution** with mean μ and variance σ^2. The special case ν_0^1 is called the **standard normal distribution**.

It is a matter of empirical observation that normal and approximately normal distributions are extremely common in applied probability and statistics. The theoretical explanation for this phenomenon is the central limit theorem, the idea of which is as follows. Suppose $\{X_j\}$ is a sequence of independent identically distributed random variables with mean 0 and variance σ^2. Then $n^{-1}\sum_1^n X_j$ has mean 0 and variance $n^{-1}\sigma^2$, so there is a high probability that it is close to 0 when n is large: this is the content of the weak law of large numbers. On the other hand, $n^{-1/2}\sum_1^n X_j$ has mean 0 and variance σ^2 for all n, so one might ask if its distribution approaches some nontrivial limit as $n \to \infty$. The remarkable answer is that no matter what the distribution of the X_j's is, this limit exists and equals the normal distribution with mean 0 and variance σ^2.

The central limit theorem is really a theorem in Fourier analysis. We shall state it as such and then translate it into probability theory.

(9.14) Theorem. Let λ be a Borel probability measure on $\mathbf{R}$ such that

$$\int t^2\, d\lambda(t) = 1, \qquad \int t\, d\lambda(t) = 0.$$

(The finiteness of the first integral implies the existence of the second.) For $n \in \mathbf{N}$ let $\lambda^{(n)} = \lambda * \cdots * \lambda$ (n factors) and define the measure λ_n by $\lambda_n(E) = \lambda^{(n)}(\sqrt{n}\,E)$, where $\sqrt{n}\,E = \{\sqrt{n}\,t \colon t \in E\}$. Then $\lambda_n \to \nu_0^1$ vaguely as $n \to \infty$.

Proof. The hypotheses on λ imply that its Fourier transform $\hat{\lambda}$ is of class C^2 and satisfies $\hat{\lambda}(0) = 1$, $\hat{\lambda}'(0) = 0$, and $\hat{\lambda}''(0) = -4\pi^2$. [See Theorem (8.22d); the same argument works for measures.] Thus by Taylor's theorem,

$$\hat{\lambda}(\xi) = 1 - 2\pi^2\xi^2 + o(\xi^2)$$

where $o(\alpha)$ denotes a quantity which satisfies $\alpha^{-1}o(\alpha) \to 0$ as $\alpha \to 0$. Moreover, $(\lambda^{(n)})\hat{} = \hat{\lambda}^n$, so by the obvious change of variable,

$$\hat{\lambda}_n(\xi) = \big[\hat{\lambda}(n^{-1/2}\xi)\big]^n = \left[1 - \frac{2\pi^2\xi^2}{n} + o\left(\frac{\xi^2}{n}\right)\right]^n.$$

Thus, since $\log(1+z) = z + o(z)$,

$$\log \hat{\lambda}_n(\xi) = n \log\left[1 - \frac{2\pi^2\xi^2}{n} + o\left(\frac{\xi^2}{n}\right)\right] = -2\pi^2\xi^2 + n \cdot o\left(\frac{\xi^2}{n}\right),$$

which tends to $-2\pi^2\xi^2$ as $n \to \infty$. In other words, $\hat{\lambda}_n(\xi) \to e^{-2\pi^2\xi^2}$ as $n \to \infty$ for all ξ, so the conclusion follows from Propositions (8.24) and (8.69). □

(9.15) The Central Limit Theorem. Let $\{X_j\}$ be a sequence of independent identically distributed L^2 random variables with mean μ and variance σ^2. Then, as $n \to \infty$, the distribution of $(\sigma\sqrt{n})^{-1}\Sigma_1^n(X_j - \mu)$ converges vaguely to the standard normal distribution ν_0^1, and

$$P\left(\frac{1}{\sigma\sqrt{n}}\sum_1^n(X_n - \mu) \le a\right) \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^a e^{-t^2/2}\,dt.$$

Proof. Replacing X_j by $\sigma^{-1}(X_j - \mu)$, we may assume that $\mu = 0$, $\sigma = 1$. If λ is the common distribution of the X_j's, then λ satisfies the hypotheses of Theorem (9.14), and in the notation used there, λ_n is the distribution of $n^{-1/2}\Sigma_1^n X_j$. The first assertion thus follows immediately, and the second one is equivalent to it by Proposition (7.19). □

As the reader may readily verify, the same argument yields the following more general result. Under the hypotheses of the central limit theorem, if $\{K_n\}$ is any sequence of finite subsets of $\mathbf{N}$ such that $k_n = \text{card}(K_n) \to \infty$, then the distribution of $(\sigma\sqrt{k_n})^{-1}\Sigma_{j\in K_n}X_j$ converges vaguely to ν_0^1.

EXERCISES

16. A fair coin is tossed 10,000 times; let X be the number of times it comes up heads. Use the central limit theorem to estimate:
- (a) the probability that $4950 \le X \le 5050$.
- (b) the number k such that $|X - 5000| \le k$ with probability 0.98. [You will need to consult a table of values of $(2\pi)^{-1/2}\int_0^x e^{-t^2/2}\,dt$.]

17. If $X_n \to X$ in probability, then $P_{X_n} \to P_X$ vaguely. [Use Proposition (7.19).]

18. Under the hypotheses of the central limit theorem, the sequence $Y_n = (\sigma\sqrt{n})^{-1}\Sigma_1^n(X_j - \mu)$ does not converge in probability. (Use the remark following the central limit theorem to show that $\{Y_{2^k}\}$ is not Cauchy in probability.)

19. If $\{X_j\}$ is a sequence of independent identically distributed random variables with mean 0 and variance 1, the distributions of

$$\frac{\sum_1^n X_j}{\sqrt{\sum_1^n X_j^2}} \quad \text{and} \quad \frac{\sqrt{n}\sum_1^n X_j}{\sum_1^n X_j^2}$$

both converge vaguely to the standard normal distribution.

20. (The moment convergence theorem) Let X_n, X be random variables such that $P_{X_n} \to P_X$ vaguely and $\sup_n E(|X_n|^r) < \infty$. Then $E(|X_n|^s) \to E(|X|^s)$ for all $s \in (0, r)$, and if also $s \in \mathbf{N}$, then $E(X_n^s) \to E(X^s)$.

21. Let $\{X_n\}$ be a sequence of independent random variables, each having the Poisson distribution with mean 1 (Exercise 8). Let

$$S_n = X_1 + \cdots + X_n, \qquad Y_n = \left(\frac{S_n - n}{\sqrt{n}}\right)^- = \max\left(\frac{n - S_n}{\sqrt{n}}, 0\right).$$

- (a) $E(Y_n) = e^{-n}\Sigma_{k=0}^n \frac{n-k}{\sqrt{n}}\frac{n^k}{k!} = \frac{n^{n+(1/2)}e^{-n}}{n!}$. (For the first equation, use Exercise 8b. As for the second, the sum telescopes.)
- (b) P_{Y_n} converges vaguely to $\frac{1}{2}\delta_0 + \chi_{(0,\infty)}\nu_0^1$. [Use Proposition (7.19).]
- (c) $\int_0^\infty t^2\,dP_{Y_n}(t) = E(Y_n^2) \le 1$ for all n.
- (d) $E(Y_n) = \int_0^\infty t\,dP_{Y_n}(t) \to \int_0^\infty t\,d\nu_0^1(t) = (2\pi)^{-1/2}$. (Use Exercise 20.) Combining (a) and (d), one obtains Stirling's formula:

$$\lim_{n\to\infty}\frac{n!}{n^n e^{-n}\sqrt{2\pi n}} = 1.$$

22. In this exercise we consider random variables with the value in the circle $\mathbf{T}^1$, regarded as $\{z \in \mathbf{C}: |z| = 1\}$. The distribution of such a random variable is a measure on $\mathbf{T}^1$.
- (a) If $X_1, \ldots, X_n$ are independent, then $P_{X_1X_2\cdots X_n} = P_{X_1} * \cdots * P_{X_n}$.
- (b) If $\{X_j\}$ is a sequence of independent random variables with a common distribution λ, the distribution of $\prod_1^n X_j$ converges vaguely to the uniform distribution on $\mathbf{T}^1$ (= arc length over 2π) unless λ is supported on a finite subgroup $Z_m = \{e^{2\pi ij/m}: 0 \le j < m\}$ of $\mathbf{T}^1$, in which case it converges to the uniform distribution $m^{-1}\Sigma_{z\in Z_m}\delta_z$ on Z_m. (Use Exercise 8.57.)

4. CONSTRUCTION OF SAMPLE SPACES

The preceding two sections have dealt with sequences of random variables whose joint distributions have certain properties. We now address the question of finding examples of such sequences, and more generally of constructing families $\{X_\alpha\}_{\alpha\in A}$ of random variables indexed by an arbitrary set A whose finite subfamilies have prescribed joint distributions.

If the index set A is finite, this is easy: given any Borel probability measure P on $\mathbf{R}^n$, P is by definition the joint distribution of the coordinate functions $X_1, \ldots, X_n$ on the space $(\mathbf{R}^n, \mathscr{B}_{\mathbf{R}^n}, P)$. If A is infinite, however, the problem is more delicate. Suppose to begin with that $\{X_\alpha\}_{\alpha\in A}$ is a family of random

variables on some sample space $(\Omega, \mathscr{B}, P)$, and for each ordered n-tuple $(\alpha_1, \dots, \alpha_n)$ of distinct elements of A $(n \in \mathbf{N})$ let $P_{(\alpha_1,\dots,\alpha_n)}$ be the joint distribution of $X_{\alpha_1}, \dots, X_{\alpha_n}$. Then the measures $P_{(\alpha_1,\dots,\alpha_n)}$ satisfy the following consistency conditions:

(9.16) If σ is a permutation of $\{1, \dots, n\}$, then $P_{(\alpha_{\sigma(1)},\dots,\alpha_{\sigma(n)})}$ is the image measure of $P_{(\alpha_1,\dots,\alpha_n)}$ under the map T_σ: $\mathbf{R}^n \to \mathbf{R}^n$ defined by $T_\sigma(x_1, \dots, x_n) = (x_{\sigma(1)}, \dots, x_{\sigma(n)})$. That is,

$$dP_{(\alpha_{\sigma(1)},\dots,\alpha_{\sigma(n)})}(x_{\sigma(1)}, \dots, x_{\sigma(n)}) = dP_{(\alpha_1,\dots,\alpha_n)}(x_1, \dots, x_n).$$

(9.17) If $k \le n$, $P_{(\alpha_1,\dots,\alpha_k)}$ is the image measure of $P_{(\alpha_1,\dots,\alpha_n)}$ under the projection π: $\mathbf{R}^n \to \mathbf{R}^k$ defined by $\pi(x_1, \dots, x_n) = (x_1, \dots, x_k)$. That is, if $E \in \mathscr{B}_{\mathbf{R}^k}$,

$$P_{(\alpha_1,\dots,\alpha_k)}(E) = P_{(\alpha_1,\dots,\alpha_n)}(E \times \mathbf{R}^{n-k}).$$

Conversely, given any family of measures $P_{(\alpha_1,\dots,\alpha_n)}$ satisfying these conditions, we shall show that there exist a sample space $(\Omega, \mathscr{B}, P)$ and random variables $\{X_\alpha\}$ on Ω such that $P_{(\alpha_1,\dots,\alpha_n)}$ is the joint distribution of $(X_{\alpha_1}, \dots, X_{\alpha_n})$. To do this, it is convenient to make one minor technical modification: we replace $\mathbf{R}$ by its one-point compactification $\mathbf{R}^* = \mathbf{R} \cup \{\infty\}$. Any Borel measure on $\mathbf{R}^n$ can be regarded as a Borel measure on $(\mathbf{R}^*)^n$ which assigns measure zero to $(\mathbf{R}^*)^n \setminus \mathbf{R}^n$, and vice versa. We shall thus allow our random variables to assume the value ∞, although they will do so with probability zero.

(9.18) Theorem. Let A be an arbitrary nonempty set, and suppose that for each ordered n-tuple of distinct elements of A $(n \in \mathbf{N})$ we are given a Borel probability measure $P_{(\alpha_1,\dots,\alpha_n)}$ on $\mathbf{R}^n$, or equivalently on $(\mathbf{R}^*)^n$, satisfying (9.16) and (9.17). Then there is a unique Radon probability measure P on the compact Hausdorff space $\Omega = (\mathbf{R}^*)^A$ such that $P_{(\alpha_1,\dots,\alpha_n)}$ is the joint distribution of $(X_{\alpha_1}, \dots, X_{\alpha_n})$, where X_α: $\Omega \to \mathbf{R}^*$ is the αth coordinate function.

Proof. Let $C_F(\Omega)$ be the set of all $f \in C(\Omega)$ which depend only on finitely many coordinates. If $f \in C_F(\Omega)$, say $f(x) = F(x_{\alpha_1}, \dots, x_{\alpha_n})$, let

$$I(f) = \int F \, dP_{(\alpha_1,\dots,\alpha_n)}.$$

$I(f)$ is well defined because of (9.16) and (9.17): if we permute the variables or add on some extra ones, we get the same answer. Clearly $I(f) \ge 0$ if $f \ge 0$ and $|I(f)| \le \|f\|_u$ with equality when f is constant.

Now, $C_F(\Omega)$ is clearly an algebra which separates points and is closed under complex conjugation, so by the Stone–Weierstrass theorem it is dense in $C(\Omega)$. The functional I thus extends uniquely to a positive linear functional on $C(\Omega)$ with norm 1, so by the Riesz representation theorem there is a unique Radon probability measure P on Ω such that $I(f) = \int f dP$ for all $f \in C(\Omega)$. Let X_α be the αth coordinate function on Ω and let $P'_{(\alpha_1,\dots,\alpha_n)}$ be the joint distribution of $(X_{\alpha_1}, \dots, X_{\alpha_n})$ on $(\mathbf{R}^*)^n$. If $F \in C((\mathbf{R}^*)^n)$ and $f = F \circ (X_{\alpha_1}, \dots, X_{\alpha_n})$ as above, then

$$\int F \, dP'_{(\alpha_1,\dots,\alpha_n)} = \int f dP = I(f) = \int F \, dP_{(\alpha_1,\dots,\alpha_n)}.$$

But $P'_{(\alpha_1,\dots,\alpha_n)}$ and $P_{(\alpha_1,\dots,\alpha_n)}$ are both Radon measures by Theorem 7.8, so they are equal by the uniqueness of the Riesz representation. □

The only property of $\mathbf{R}^*$ used in this proof is the fact that it is a compact Hausdorff space; evidently the same argument yields an analogous existence theorem for random variables with values in an arbitrary compact Hausdorff space, assuming the measures $P_{(\alpha_1,\dots,\alpha_n)}$ are Radon. In particular, if there is a compact set $K \subset \mathbf{R}$ such that $P_{(\alpha_1,\dots,\alpha_n)}$ is supported in K^n for all $(\alpha_1, \dots, \alpha_n)$, we could replace $\mathbf{R}^*$ by K and thus avoid introducing the point at infinity.

Of special interest is the independent case. That is, suppose $\{P_\alpha\}_{\alpha \in A}$ is a family of probability measures on $\mathbf{R}$. If we define $P_{(\alpha_1,\dots,\alpha_n)} = \prod_1^n P_{\alpha_j}$, the conditions (9.16) and (9.17) are automatically satisfied, and the resulting measure P on $(\mathbf{R}^*)^A$ is, in an obvious sense, the product of the P_α's. More generally, we have:

(9.19) Theorem. Let $\{\Omega_\alpha\}_{\alpha \in A}$ be a family of compact Hausdorff spaces, and for each α let P_α be a Radon probability measure on Ω_α. Then there is a unique Radon probability measure P on $\Omega = \prod_{\alpha \in A} \Omega_\alpha$ such that for any distinct $\alpha_1, \dots, \alpha_n \in A$ the image measure of P on $\prod_1^n \Omega_{\alpha_j}$ under the natural projection is the Radon product of $P_{\alpha_1}, \dots, P_{\alpha_n}$.

The proof is identical to that of Theorem (9.18).

EXERCISES

The following exercises form a connected set.

23. Given $b \in \mathbf{N} \setminus \{1\}$, let $B = \{0, 1, \dots, b-1\}$ and $\Omega = B^{\mathbf{N}}$. Put the discrete topology on B and the product topology on Ω, and let P be the product measure on Ω given by Theorem (9.19), where each P_n is b^{-1} times counting measure on B. Let $\{X_n\}_1^\infty$ be the coordinate functions on Ω. Then if $A_1, \dots, A_n \subset B$,

$$P\left(\bigcap_1^n X_j^{-1}(A_j)\right) = b^{-n} \prod_1^n \operatorname{card}(A_j),$$

and $P(\{\omega\}) = 0$ for all $\omega \in \Omega$.

24. Let $\Omega' = \{\omega \in \Omega\colon X_n(\omega) \neq 0 \text{ for infinitely many } n\}$. Then $\Omega \setminus \Omega'$ is countable and $P(\Omega') = 1$. Define $F\colon \Omega \to [0,1]$ by $F(\omega) = \sum_1^\infty X_n(\omega) b^{-n}$ [so $F(\omega)$ is the number such that $\{X_n(\omega)\}$ is the sequence of digits in its base b decimal expansion]. Then $F|\Omega'$ is a bijection from Ω' to $(0,1]$ which maps $\mathscr{B}_{\Omega'}$ bijectively onto $\mathscr{B}_{(0,1]}$. ($\mathscr{B}_{\Omega'}$ is generated by sets of the form $X_n^{-1}(A) \cap \Omega'$. The image under F of such a set is a finite union of intervals of the form $(jb^{-n}, kb^{-n}]$, and these sets generate $\mathscr{B}_{(0,1]}$.)

25. The image measure of P under F is Lebesgue measure.

26. (Borel's normal number theorem) A number $x \in (0,1]$ is called **normal** in base b if the digits $0, 1, \ldots, b-1$ occur with equal frequency in its base b decimal expansion, that is, if $n^{-1} X_j^{-1}(F^{-1}(x)) \to b^{-1}$ as $n \to \infty$. Almost every $x \in (0,1]$ (with respect to Lebesgue measure) is normal in base b for every b.

5. THE WIENER PROCESS

It is observed that small particles suspended in a fluid such as water or air undergo an irregular motion, known as Brownian motion, due to the collisions of the particles with the molecules of the fluid. A physical theory of this phenomenon, in the limiting case where the motion is assumed to result from an infinite number of collisions with molecules of infinitesimal size, was developed independently by Einstein and Smoluchowski, and it was later put on a sound mathematical basis by Wiener. The resulting mathematical model, called the Wiener process or Brownian motion process, has since turned out to be of central importance in probability theory and its applications to physics and mathematical analysis.

One can consider Brownian motion in any number of space dimensions. We shall describe the theory in dimension one and then indicate how to generalize it. The position of a particle undergoing Brownian motion on the line at time $t > 0$ is to be considered as a random variable X_t (on a sample space to be specified later) satisfying the following conditions. First, as a matter of normalization, we assume that the particle starts at the origin at time $t = 0$:

(9.20) $X_0 = 0$ (almost surely).

Second, since any given collision affects the particle only by an infinitesimal amount, it has no long-term effect, so the motion of the particle after time t should depend on its position X_t at that time but not on its previous history. Thus we assume:

(9.21) If $0 \leq t_0 < t_1 < \cdots < t_n$, the random variables $X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.

Since the physical processes underlying Brownian motion are homogeneous in time, it is reasonable to assume that the distribution of $X_t - X_s$ depends only

on $t - s$. If we divide the interval $[s,t]$ into n equal subintervals and write $X_t - X_s = \sum_1^n (X_{t_j} - X_{t_{j-1}})$ where $t_j = s + j(t-s)/n$, we therefore see that $X_t - X_s$ is a sum of independent identically distributed random variables. Since n may be taken arbitrarily large, the central limit theorem suggests that the distribution of $X_t - X_s$ should be normal. This is only a plausibility argument, but the conclusion is correct on physical grounds: $X_t - X_s$ is normally distributed. By spatial symmetry (the particle is as likely to move to the left as to the right), its expected value is zero. Moreover, the subdivision argument given above, together with Corollary (9.6), leads to the conclusion that $\sigma^2(X_t - X_s) = (t-s)\sigma^2(X_1 - X_0)$ whenever $t - s$ is positive and rational. Making the natural assumption that $\sigma^2(X_t - X_s)$ is continuous in $t - s$, we are lead to the third condition:

(9.22) There is a constant $C > 0$ such that for $0 < t < s$, the distribution of $X_t - X_s$ is the normal distribution $\nu_0^{C(t-s)}$ with mean 0 and variance $C(t-s)$.

The constant C, which expresses the rate of diffusion, is of course related to the physical parameters of the system. For simplicity, we shall henceforth take $C = 1$.

A family $\{X_t\}_{t \geq 0}$ of random variables satisfying (9.20)–(9.22) (with $C = 1$) is called an **abstract Wiener process**. The generalization to n dimensions can now be described easily: an ***n*-dimensional abstract Wiener process** is a family of $\mathbf{R}^n$-valued random variables $\{\boldsymbol{X}_t\}_{t \geq 0}$, where $\boldsymbol{X}_t = (X_t^1, \ldots, X_t^n)$, such that (a) $\{X_t^j\}_{t \geq 0}$ is a one-dimensional abstract Wiener process for each j, and (b) if $0 \leq t_1 < \cdots < t_K$, the random variables $X_{t_k}^j - X_{t_{k-1}}^j$ $(1 \leq j \leq n, 1 \leq k \leq K)$ are independent. This implies that for $t > s$, $\boldsymbol{X}_t - \boldsymbol{X}_s$ has the n-dimensional normal distribution:

$$d\left[\nu_0^{t-s}\right]^n(x_1, \ldots, x_n) = \left[2\pi(t-s)\right]^{-n/2} \exp\left(\sum_1^n x_j^2 / 2(t-s)\right) dx_1 \cdots dx_n,$$

which has the appropriate sort of spherical symmetry.

We return to the one-dimensional case. The conditions (9.20)–(9.22) completely determine the joint distribution of the X_t's, as follows. If $t_1 < \cdots < t_n$, then $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent (since $X_{t_1} = X_{t_1} - X_0$ a.s.), so their joint distribution is the product measure

$$\nu_0^{t_1} \times \nu_0^{t_2 - t_1} \times \cdots \times \nu_0^{t_n - t_{n-1}}.$$

But

$$(X_{t_1}, \ldots, X_{t_n}) = T(X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}})$$

where

$$T(y_1, \ldots, y_n) = (y_1, y_1 + y_2, \ldots, y_1 + \cdots + y_n).$$

Since $\det T = 1$, Theorem (2.43) implies that the joint distribution $P_{(t_1, \ldots, t_n)}$ of

$X_{t_1}, \dots, X_{t_n}$ is given by

(9.23)

$$\int f\,dP_{(t_1,\dots,t_n)} = \int f(T(y))\,d\left(\nu_0^{t_1} \times \nu_0^{t_2-t_1} \times \cdots \times \nu_0^{t_n-t_{n-1}}\right)(y)$$

$$= \left[\prod_1^n 2\pi(t_j - t_{j-1})\right]^{-1/2} \int \cdots \int f(T(y_1,\dots,y_n))$$

$$\times \exp\left(\frac{\sum_1^n y_j^2}{2(t_j - t_{j-1})}\right) dy_1 \cdots dy_n$$

$$= \left[\prod_1^n 2\pi(t_j - t_{j-1})\right]^{-1/2} \int \cdots \int f(x_1,\dots,x_n)$$

$$\times \exp\left(\frac{\sum_1^n (x_j - x_{j-1})^2}{2(t_j - t_{j-1})}\right) dx_1 \cdots dx_n$$

$$= \int \cdots \int f(x_1,\dots,x_n)\,d\nu_{x_{n-1}}^{t_n - t_{n-1}}(x_n) \cdots d\nu_{x_1}^{t_2-t_1}(x_2)\,d\nu_0^{t_1}(x_1),$$

where, in the intermediate steps, we have set $t_0 = 0$ and $x = Ty$. We thus know $P_{(t_1,\dots,t_n)}$ when $t_1 < \cdots < t_n$, and we then obtain it in the general case by permuting the variables according to (9.16). Also, it follows easily from (9.23) that (9.17) is satisfied. Therefore, by Theorem (9.18), abstract Wiener processes exist.

This situation leaves something to be desired, however. Physically, one expects the position of a particle to be a continuous function of time, so one would like the sample space for the Wiener process to be $C([0,\infty), \mathbf{R})$ (or some subset thereof) and the random variable X_t to be evaluation at t. Actually, Theorem (9.18) yields something along these lines: the sample space it provides is the space $(\mathbf{R}^*)^{[0,\infty)}$ of *all* functions from $[0,\infty)$ into the compactified line, and $X_t(\omega)$ is indeed $\omega(t)$. We can therefore achieve our goal by showing that the measure P of Theorem (9.18) is concentrated on $C([0,\infty), \mathbf{R})$ [considered as a subset of $(\mathbf{R}^*)^{[0,\infty)}$]. The resulting realization of the abstract Wiener process on $C([0,\infty), \mathbf{R})$ is what is usually called the **Wiener process**.

(9.24) Theorem. Let $\Omega = (\mathbf{R}^*)^{[0,\infty)}$, $\Omega_c = C([0,\infty), \mathbf{R})$, and P the Radon measure on Ω whose finite-dimensional distributions are given by (9.23). Then Ω_c is a Borel subset of Ω and $P(\Omega_c) = 1$.

Proof. To begin with, we need to make a few comments about the role of the point at infinity. The function $f(t,s) = |t - s|$ maps $\mathbf{R}^2$ to $[0, +\infty)$ (we write $+\infty$ to distinguish it from the point at infinity in $\mathbf{R}^*$), and we extend f to a map from $(\mathbf{R}^*)^2$ to $[0, +\infty]$ by declaring that $|t - \infty| = |\infty - t| = +\infty$ for $t \in \mathbf{R}$, and $|\infty - \infty| = 0$. When thus extended, f is of course discontinuous at ∞, but it is lower semicontinuous, as the reader may verify (Exercise 27). Thus for $a \in (0,\infty)$ the sets $\{\omega \in \Omega\colon |\omega(t) - \omega(s)| > a\}$ are open and the sets $\{\omega \in \Omega\colon |\omega(t) - \omega(s)| \le a\}$ are closed.

Now, a real-valued function ω on $[0,\infty)$ is continuous iff it is uniformly continuous on $[0,n]$ for each $n \in \mathbf{N}$, and it is uniformly continuous on $[0,n]$ iff for every $j \in \mathbf{N}$ there exists $k \in \mathbf{N}$ such that $|\omega(t) - \omega(s)| \le j^{-1}$ [note the use of $\le$ rather than $<$] for all $s \in [0,n]$ and all $t \in [0,n] \cap (s - k^{-1}, s + k^{-1})$. Moreover, even if we only assume that ω is $\mathbf{R}^*$-valued, this last condition implies that it is real-valued unless it is identically ∞. Therefore, if we define $\omega_\infty(t) = \infty$ for all t, we have

$$\Omega_c \cup \{\omega_\infty\} = \bigcap_{n=1}^{\infty} \bigcap_{j=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{s\in[0,n]} \bigcap_{|t-s|<1/k} \{\omega \in \Omega\colon |\omega(t) - \omega(s)| \le j^{-1}\}.$$

By the preceding remarks, $\{\omega\colon |\omega(t) - \omega(s)| \le j^{-1}\}$ is closed for all s, t, and j. Hence $\Omega_c \cup \{\omega_\infty\}$ is an $F_{\sigma\delta}$ set, and Ω_c is therefore a Borel set.

Moreover, if for $\varepsilon, \delta > 0$ and $n \in \mathbf{N}$ we set

$$U(n,\varepsilon,\delta) = \{\omega \in \Omega\colon |\omega(t) - \omega(s)| > 8\varepsilon \text{ for some } t,s \in [0,n] \text{ with } |s - t| < \delta\},$$

we have

$$(\Omega_c)^c = \left[\bigcup_{n=1}^{\infty} \bigcup_{j=1}^{\infty} \bigcap_{k=1}^{\infty} U\left(n, (8j)^{-1}, k^{-1}\right)\right] \cup \{\omega_\infty\}.$$

Clearly $P(\{\omega_\infty\}) = 0$, so in order to show that $P(\Omega_c) = 1$, or equivalently that $P((\Omega_c)^c) = 0$, it will suffice to show that $\lim_{k\to\infty} P(U(n,\varepsilon,k^{-1})) = 0$ for all ε and n. To achieve this we need to make some estimates in terms of the quantity

$$\rho(\varepsilon,\delta) = \sup_{t\le\delta} \int_{|x|>\varepsilon} d\nu_0^t(x) = \sup_{t\le\delta} \left(\frac{2}{\pi t}\right)^{1/2} \int_\varepsilon^\infty e^{-x^2/2t}\,dx.$$

(9.25) Lemma. For each $\varepsilon > 0$, $\lim_{\delta\to 0} \delta^{-1}\rho(\varepsilon,\delta) = 0$.

Proof. The following crude estimate does the job: if $\delta < \varepsilon^2$,

$$\rho(\varepsilon,\delta) \le \sup_{t\le\delta} \left(\frac{2}{\pi t}\right)^{1/2} \int_\varepsilon^\infty e^{-\varepsilon x/2t}\,dx = \frac{2}{\varepsilon}\left(\frac{2}{\pi\delta}\right)^{1/2} e^{-\varepsilon^2/2\delta}. \qquad \square$$

(9.26) Lemma. Suppose $\varepsilon > 0$, $\delta > 0$, $0 \le t_1 < \cdots < t_k$, $t_k - t_1 \le \delta$, and

$$A = \{\omega: |\omega(t_j) - \omega(t_1)| > 2\varepsilon \text{ for some } j \in \{2, \ldots, k\}\}.$$

Then $P(A) \le 2\rho(\varepsilon, \delta)$.

Proof. Let

$$B_j = \{\omega: |\omega(t_k) - \omega(t_j)| > \varepsilon\} \quad (j = 1, \ldots, k),$$

$$D_j = \{\omega: |\omega(t_j) - \omega(t_1)| > 2\varepsilon \text{ and } |\omega(t_i) - \omega(t_1)| \le 2\varepsilon \text{ for } i < j\}$$

$$(j = 2, \ldots, k).$$

If $\omega \in A$, then $\omega \in D_j$ for some $j \ge 2$; if ω is not also in B_j, then ω must be in B_1, for $|\omega(t_k) - \omega(t_1)| > \varepsilon$ whenever $|\omega(t_j) - \omega(t_1)| > 2\varepsilon$ and $|\omega(t_k) - \omega(t_j)| \le \varepsilon$. In other words,

$$A \subset B_1 \cup \bigcup_2^n (B_j \cap D_j).$$

But $P(B_j) \le \rho(\varepsilon, \delta)$ for each j by (9.22), and B_j and D_j are independent events by (9.21); also, the D_j's are clearly disjoint. Therefore,

$$P(A) \le P(B_1) + \sum_2^n P(B_j)P(D_j) \le \rho(\varepsilon, \delta)\left[1 + \sum_2^n P(D_j)\right] \le 2\rho(\varepsilon, \delta).$$

□

(9.27) Lemma. With the notation of Lemma (9.26), let

$$E = \{\omega: |\omega(t_i) - \omega(t_j)| > 4\varepsilon \text{ for some } i, j \in \{1, \ldots, k\}\}.$$

Then $P(E) \le 2\rho(\varepsilon, \delta)$.

Proof. If $|\omega(t_i) - \omega(t_j)| > 4\varepsilon$, then either $|\omega(t_i) - \omega(t_1)| > 2\varepsilon$ or $|\omega(t_j) - \omega(t_1)| > 2\varepsilon$. Thus $E \subset A$, so the result follows from Lemma (9.26). □

(9.28) Lemma. Suppose $\varepsilon > 0$, $0 \le a < b$, and $b - a \le \delta$. If

$$V = \{\omega: |\omega(t) - \omega(s)| > 4\varepsilon \text{ for some } t, s \in [a, b]\}$$

then $P(V) \le 2\rho(\varepsilon, \delta)$.

Proof. If S is a finite subset of $[a, b]$, let

$$V(S) = \{\omega: |\omega(t) - \omega(s)| > 4\varepsilon \text{ for some } t, s \in S\}.$$

Then the sets $V(S)$ are open in Ω and their union is V, and by Lemma (9.27),

$P(V(S)) \le 2\rho(\varepsilon, \delta)$. Since the family $\{V(S): S \text{ is a finite subset of } [a, b]\}$ is closed under finite unions, if K is any compact subset of V we have $K \subset V(S)$ for some S and hence $P(K) \le 2\rho(\varepsilon, \delta)$. But then $P(V) \le 2\rho(\varepsilon, \delta)$ by the inner regularity of P. □

We can now finish the proof of Theorem (9.24). Given $\varepsilon > 0$ and $n, k \in \mathbf{N}$, divide the interval $[0, n]$ into the nk subintervals $[0, k^{-1}]$, $[k^{-1}, 2k^{-1}], \ldots, [n - k^{-1}, n]$. If $\omega \in U(n, \varepsilon, k^{-1})$, then $|\omega(t) - \omega(s)| > 8\varepsilon$ for some t, s lying in the same or adjacent subintervals, and hence, in the notation of Lemma (9.28), $\omega \in V$ where $\delta = k^{-1}$ and $[a, b]$ is one of the subintervals. By Lemmas (9.25) and (9.28) we therefore have

$$P(U(n, \varepsilon, k^{-1})) \le 2nk\rho(\varepsilon, k^{-1}) \to 0 \text{ as } k \to \infty,$$

which completes the proof. □

EXERCISES

27. The function $f: (\mathbf{R}^*)^2 \to [0, +\infty]$ defined by $f(t, s) = |t - s|$ for $t, s \in \mathbf{R}$, $f(\infty, t) = f(t, \infty) = +\infty$ for $t \in \mathbf{R}$, and $f(\infty, \infty) = 0$ is lower semicontinuous.

28. If one replaces (9.20) by the more general assumption that the distribution of X_0 is a given probability measure on $\mathbf{R}$, the results of this section remain valid, with a few obvious modifications. Discuss.

29. Let $\Omega = (\mathbf{R}^*)^{[0,\infty)}$, $\{X_t\}_{t \ge 0}$ the coordinate functions on Ω, and for any $A \subset [0, \infty)$, $\mathcal{M}_A$ = the σ-algebra generated by $\{X_t\}_{t \in A}$. (Thus $\mathcal{M}_{[0,\infty)}$ is the product σ-algebra on Ω corresponding to the Borel σ-algebras on the factors.)

(a) If $V \in \mathcal{M}_A$, $\omega \in V$, and $\omega'(t) = \omega(t)$ for all $t \in A$, then $\omega' \in V$.

(b) If $V \in \mathcal{M}_{[0,\infty)}$, then $V \in \mathcal{M}_A$ for some countable set A.

(c) The space $\Omega_c = C([0, \infty), \mathbf{R})$ is not in $\mathcal{M}_{[0,\infty)}$.

30. If $\omega \in \Omega_c = C([0, \infty), \mathbf{R})$, it can be shown that ω is almost surely not of bounded variation, so if f is a Borel measurable function on $[0, \infty)$ the integral $\int_0^\infty f(t)\, d\omega(t)$ apparently makes no sense. However:

(a) If $f = \sum_1^n c_j \chi_{[a_j, b_j)}$ is a step function, define

$$I_f(\omega) = \int_0^\infty f(t)\, d\omega(t) = \sum_1^n c_j[\omega(b_j) - \omega(a_j)].$$

Then I_f is an L^2 random variable on Ω_c with mean 0 and variance $\|f\|_2^2 = \int |f|^2\, dx$. [Hint: the intervals $[a_j, b_j)$ may be assumed disjoint.]

(b) The map $f \to I_f$ extends to an isometry from $L^2([0, \infty))$ to $L^2(\Omega_c)$.

(c) If $f \in BV([0, \infty))$ is right continuous and $\text{supp}(f)$ is compact, there is a sequence $\{f_n\}$ of step functions such that $f_n \to f$ in L^2 and $df_n \to df$ vaguely, where df denotes the Lebesgue–Stieltjes measure defined by f. [By Exercise 8.59 there is a sequence $\{\mu_n\}$ of linear combinations of point masses such that $d\mu_n \to df$ vaguely. Consider $f_n(x) = \mu_n((0, x]) + f(0)$.]

(d) If $f \in BV([0, \infty))$ is right continuous and $\text{supp}(f)$ is compact, then $I_f(\omega) = -\int_0^\infty \omega(t) df(t)$ almost surely. [Check this directly when f is a step function and apply (b) and (c).]

6. NOTES AND REFERENCES

The mathematical study of probability began some two and a half centuries ago, but until the advent of modern analytic tools the theory was limited to combinatorial theorems involving discrete sample spaces and a few more general results of somewhat doubtful rigorousness. In the 1920s the subject began to develop more rapidly with the aid of the Lebesgue–Stieltjes integral, and in 1933 Kolmogorov [59] put it on a solid foundation by explicitly identifying sample spaces and random variables with measure spaces and measurable functions. Since then it has grown extensively.

More detailed accounts of probability theory on a level comparable to that of this book can be found in Billingsley [12], Chung [17], and Lamperti [61].

Section 3. An account of the long history of the central limit theorem can be found in Adams [1]. More general versions of this result exist in which the random variables are not assumed to be identically distributed; see the references given above. The proof of Stirling's formula outlined in Exercise 21 is due to Wong [111].

Section 4. Theorem (9.18) is a variant due to Nelson [70] of the fundamental existence theorem of Kolmogorov [59]. In Kolmogorov's original construction the sample space is $\mathbf{R}^A$ [which could be replaced by $(\mathbf{R}^*)^A$] and the σ-algebra on which the measure P is constructed is $\otimes_{\alpha \in A} \mathscr{B}_{\mathbf{R}}$. Theorem (9.18) is a decided improvement on Kolmogorov's theorem, both in the simplicity of its proof and in the fact that the Borel σ-algebra on $(\mathbf{R}^*)^A$ properly includes $\otimes_{\alpha \in A} \mathscr{B}_{\mathbf{R}^*}$ when A is uncountable. (The significance of the latter fact is apparent from Exercise 29.)

Theorem (9.18) cannot be divorced completely from topological considerations; even in Kolmogorov's argument the topology of $\mathbf{R}$ is used in an essential way. However, there is a purely measure-theoretic version of Theorem (9.19). Namely, if $\{(\Omega_\alpha, \mathscr{B}_\alpha, P_\alpha)\}_{\alpha \in A}$ is any family of probability spaces, there is a unique product measure P on $(\prod_{\alpha \in A} \Omega_\alpha, \otimes_{\alpha \in A} \mathscr{B}_\alpha)$ whose finite-dimensional

projections are the finite products of the P_α's. See Hewitt and Stromberg [54], Section 22, or Halmos [41], Section 38.

Section 5. Wiener developed his ideas on Brownian motion in a series of papers in the early 1920s; his construction of the measure P on $C([0, \infty), \mathbf{R})$ is quite different from ours. Our proof of Theorem (9.24) follows Nelson [71]. See also Nelson [72] for some related material, including the derivation of the postulates (9.20)–(9.22) from physical principles.

A discussion of the many interesting properties of the Wiener process is beyond the scope of this book. We shall mention only one, as a complement to Theorem (9.24): the sample paths of the Wiener process are almost surely nowhere differentiable. In fact, with probability one, at each point they are Hölder continuous of every exponent $\alpha < \frac{1}{2}$ but not of exponent $\frac{1}{2}$. This fact, which seems startling at first, looks more reasonable when one reflects that $[\omega(t) - \omega(s)]/|t - s|^{1/2}$ has the standard normal distribution for all t, s.

10

MORE MEASURES AND INTEGRALS

In this chapter we discuss some further examples of measures and integrals which are of importance in analysis and geometry: invariant measures on locally compact groups, geometric measures of lower-dimensional sets in $\mathbf{R}^n$, and integration of densities and differential forms on manifolds. Although we have grouped these topics together in one chapter, they are substantially independent of one another.

1. TOPOLOGICAL GROUPS AND HAAR MEASURE

A **topological group** is a group G endowed with a topology such that the group operations $(x, y) \to xy$ and $x \to x^{-1}$ are continuous from $G \times G$ and G to G. Examples include topological vector spaces (the group operation being addition), groups of invertible $n \times n$ real matrices (with the relative topology induced from $\mathbf{R}^{n^2}$), and all groups equipped with the discrete topology. If G is a topological group, we denote the identity element of G by e, and if $A, B \subset G$ and $x \in G$ we define

$$xA = \{xy\colon y \in A\}, \qquad Ax = \{yx\colon y \in A\},$$

$$A^{-1} = \{x^{-1}\colon x \in A\}, \qquad AB = \{yz\colon y \in A, z \in B\}.$$

We say that $A \subset G$ is **symmetric** if $A = A^{-1}$. Here are some of the basic properties of topological groups:

(10.1) Proposition. Let G be a topological group.

(a) The topology of G is translation invariant: if U is open and $x \in G$, then xU and Ux are open.

(b) For every neighborhood U of e there is a symmetric neighborhood V of e with $V \subset U$.

(c) For every neighborhood U of e there is a neighborhood V of e with $VV \subset U$.

(d) If H is a subgroup of G, then so is its closure $\overline{H}$.

(e) Every open subgroup of G is also closed.

(f) If K_1, K_2 are compact subsets of G, so is K_1K_2.

Proof. (a) is equivalent to the continuity in each variable of the map $(x, y) \to xy$, and (b) and (c) are equivalent to the continuity of $x \to x^{-1}$ and $(x, y) \to xy$ at the identity. For (d), if $x, y \in \overline{H}$ there exist nets $\langle x_\alpha \rangle_{\alpha \in A}$, $\langle y_\beta \rangle_{\beta \in B}$ in H which converge to x and y. Then $x_\alpha^{-1} \to x^{-1}$ and $x_\alpha y_\beta \to xy$ (where $A \times B$ is directed as a product space), so x^{-1} and xy belong to $\overline{H}$. For (e), if H is an open subgroup the cosets xH are open for all x, so that $G \setminus H = \bigcup_{x \notin H} xH$ is open and H is closed. Finally, (f) is true because K_1K_2 is the image of the compact set $K_1 \times K_2$ under the continuous map $(x, y) \to xy$. □

If f is a continuous function on the topological group G and $y \in G$, we define the left and right translates of f through y by

$$L_y f(x) = f(y^{-1}x), \qquad R_y f(x) = f(xy).$$

(The point of using y^{-1} on the left and y on the right is to make $L_{yz} = L_y L_z$ and $R_{yz} = R_y R_z$.) f is called **left** (resp. **right**) **uniformly continuous** if for every $\varepsilon > 0$ there is a neighborhood V of e such that $\|L_y f - f\|_u < \varepsilon$ (resp. $\|R_y f - f\|_u < \varepsilon$) for $y \in V$. (Some authors reverse the roles of L_y and R_y in this definition.)

(10.2) Proposition. If $f \in C_c(G)$, then f is left and right uniformly continuous.

Proof. We shall consider right uniform continuity; the proof on the left is the same. Let $K = \operatorname{supp}(f)$ and suppose $\varepsilon > 0$. For each $x \in K$ there is a neighborhood U_x of e such that $|f(xy) - f(x)| < \varepsilon/2$ for $y \in U_x$, and by Proposition (10.1b, c) there is a symmetric neighborhood V_x of e such that $V_x V_x \subset U_x$. Then $\{xV_x\}_{x \in K}$ covers K, so there exist $x_1, \ldots, x_n \in K$ such that $K \subset \bigcup_1^n x_j V_{x_j}$. Let $V = \bigcap_1^n V_{x_j}$. If $x \in K$, so that $x_j^{-1}x \in V_{x_j}$ for some j, and $y \in V$, then

$$|f(xy) - f(x)| \le |f(xy) - f(x_j)| + |f(x_j) - f(x)| < \varepsilon$$

since $xy = x_j(x_j^{-1}y) \in x_j U_{x_j}$. If $x \notin K$ and $y \in V$ then $f(x) = 0$, and either $f(xy) = 0$ or $x_j^{-1}xy \in V_{x_j}$ for some j; in the latter case $x_j^{-1}x = x_j^{-1}xyy^{-1} \in U_{x_j}$, so that $|f(x_j)| < \varepsilon/2$ and hence $|f(xy)| < \varepsilon$. □

One usually assumes that the topology of a topological group is Hausdorff. The following proposition shows that this is not much of a restriction.

(10.3) Proposition. Let G be a topological group.

(a) If G is T_1, then G is Hausdorff.

(b) If G is not T_1, let H be the closure of $\{e\}$. Then H is a normal subgroup, and if G/H is given the quotient topology (i.e., a set in G/H is open iff its inverse image in G is open), G/H is a Hausdorff topological group.

Proof. (a) If G is T_1 and $x \neq y \in G$, by Proposition (10.1b, c) there is a symmetric neighborhood V of e such that $xy^{-1} \notin VV$. Then Vx and Vy are disjoint neighborhoods of x and y, for if $z = v_1 x = v_2 y$ for some $v_1, v_2 \in V$, then $xy^{-1} = v_1^{-1} z z^{-1} v_2 \in V^{-1}V = VV$.

(b) H is a subgroup by Proposition (10.1d); it is clearly the smallest closed subgroup of G. It follows that H is normal, for if H' were a conjugate of H with $H' \neq H$, $H' \cap H$ would be a smaller closed subgroup. It is routine to verify that the group operations on G/H are continuous in the quotient topology, so that G/H is a topological group. If $\bar{e}$ is the identity element of G/H, then $\{\bar{e}\}$ is closed since its inverse image in G is H. But then every singleton set in G/H is closed by translation invariance, so G/H is T_1 and hence Hausdorff. □

In the context of Proposition (10.3), it is easy to see that every Borel measurable function on G is constant on the cosets of H and hence is effectively a function on G/H. Thus for most purposes one may as well work in the Hausdorff group G/H. We shall be interested in the case where G is locally compact, and we henceforth use the term **locally compact group** to mean a topological group whose topology is locally compact and Hausdorff.

If G is a locally compact group, a **left** (resp. **right**) **Haar measure** on G is a nonzero Radon measure μ on G which satisfies $\mu(xE) = \mu(E)$ [resp. $\mu(Ex) = \mu(E)$] for all $x \in G$ and $E \in \mathscr{B}_G$. For example, Lebesgue measure is a (left and right) Haar measure on $\mathbf{R}^n$, and counting measure is a (left and right) Haar measure on any group with the discrete topology. (Other examples will be found in the exercises.) The following proposition summarizes some elementary properties of Haar measures; in it, and in the sequel, we employ the notation

$$C_c^+ = \{f \in C_c(G): f \geq 0 \text{ and } \|f\|_u > 0\}.$$

(10.4) Proposition. Let G be a locally compact group.

(a) A Radon measure μ on G is a left Haar measure iff the measure $\tilde{\mu}$ defined by $\tilde{\mu}(E) = \mu(E^{-1})$ is a right Haar measure.

(b) A nonzero Radon measure μ on G is a left Haar measure iff $\int f d\mu = \int L_y f d\mu$ for all $f \in C_c^+$ and $y \in G$.

(c) If μ is a left Haar measure on G, then $\mu(U) > 0$ for all nonempty open $U \subset G$, and $\int f d\mu > 0$ for all $f \in C_c^+$.

(d) If μ is a left Haar measure on G, then $\mu(G) < \infty$ iff G is compact.

Proof. (a) is obvious. The "only if" implication of (b) follows from approximating f by simple functions, and the converse is true by (7.3). As for (c), since $\mu \neq 0$, by the regularity of μ there is a compact $K \subset G$ with $\mu(K) > 0$. If U is open and nonempty, K can be covered by finitely many left translates of U, and it follows that $\mu(U) > 0$. If $f \in C_c^+$, let $U = \{x: f(x) > \frac{1}{2}\|f\|_u\}$. Then $\int f d\mu \geq \frac{1}{2}\|f\|_u \mu(U) > 0$.

Finally, we prove (d). If G is compact, then $\mu(G) < \infty$ since μ is Radon. If G is not compact and V is a compact neighborhood of e, then G cannot be covered by finitely many translates of V, so there is a sequence $\{x_n\}$ such that $x_n \notin \bigcup_1^{n-1} x_i V$ for all n. Let U be a neighborhood of e such that $UU^{-1} \subset V$. If $m > n$ and $x_n U \cap x_m U$ is nonempty, then $x_m \in x_n UU^{-1} \subset x_n V$, a contradiction. Hence $\{x_m U\}_1^\infty$ is a disjoint sequence, and $\mu(x_m U) = \mu(U) > 0$ by (c), whence $\mu(G) \geq \mu(\bigcup_1^\infty x_m U) = \infty$. □

In view of Proposition (10.4a), one can pass from left Haar measures to right Haar measures at will, so for the sake of definiteness we shall concentrate our attention on left Haar measures. The following theorem is the principal result of this section.

(10.5) Theorem. Every locally compact group G possesses a left Haar measure.

Before proceeding to the formal proof, we give some motivation. If $E \in \mathscr{B}_G$ and V is open and nonempty, let $(E:V)$ denote the smallest number of left translates of V that cover E, that is,

$$(E:V) = \inf\left\{\#(A): E \subset \bigcup_{x \in A} xV\right\}$$

where $\#(A) = \text{card}(A)$ if A is finite and $\#(A) = \infty$ otherwise. $(E:V)$ is thus a rough measure of the relative sizes of E and V. If we fix a precompact open set E_0, the ratio $(E:V)/(E_0:V)$ gives a rough estimate of the size of E when the size of E_0 is normalized to be 1. This estimate becomes the more accurate the smaller V is, and it is obviously left-invariant as a function of E. We might therefore hope to obtain a Haar measure as a limit of the "quasi-measures" $E \to (E:V)/(E_0:V)$ as V shrinks to $\{e\}$.

This idea can be made to work as it stands, but it is simpler to carry out if we think of integrals of functions instead of measures of sets. If $f, \varphi \in C_c^+$, then $\{x: \varphi(x) > \frac{1}{2}\|\varphi\|_u\}$ is open and nonempty, so finitely many left translates of it cover $\text{supp}(f)$, and it follows that

$$f \leq \frac{2\|f\|_u}{\|\varphi\|_u} \sum_1^n L_{x_j} \varphi \text{ for some } x_1, \ldots, x_n \in G.$$

It therefore makes sense to define the "Haar covering number" of f with

respect to φ:

$$(f:\varphi) = \inf\left\{\sum_1^n c_j : f \le \sum_1^n c_j L_{x_j}\varphi \text{ for some } n \in \mathbf{N}, x_1, \dots, x_n \in G\right\}.$$

The quantity $(f:\varphi)$ has the following properties:

(10.6) $(f:\varphi) = (L_x f:\varphi)$ for any $x \in G$.

(10.7) $(f_1 + f_2:\varphi) \le (f_1:\varphi) + (f_2:\varphi)$.

(10.8) $(cf:\varphi) = c(f:\varphi)$ for any $c > 0$.

(10.9) $(f_1:\varphi) \le (f_2:\varphi)$ whenever $f_1 \le f_2$.

(10.10) $(f:\varphi) \ge \|f\|_u/\|\varphi\|_u$.

(10.11) $(f:\varphi) \le (f:g)(g:\varphi)$ for any $g \in C_c^+$.

The first five of these are obvious, and (10.11) follows from the observation that if $f \le \Sigma c_i L_{x_i} g$ and $g \le \Sigma d_j L_{y_j}\varphi$ then $f \le \Sigma_{i,j} c_i d_j L_{x_i y_j}\varphi$, so that

$$(f:\varphi) \le \inf \sum_{i,j} c_i d_j = \left(\inf \sum c_i\right)\left(\inf \sum d_j\right) = (f:g)(g:\varphi).$$

At this point we make a normalization by choosing $f_0 \in C_c^+$ once and for all and defining

$$I_\varphi(f) = \frac{(f:\varphi)}{(f_0:\varphi)} \text{ for } f, \varphi \in C_c^+.$$

For fixed φ, then, the functional I_φ is left-invariant and bears some resemblance to a positive linear functional except that it is only subadditive. Moreover, by (10.11) it satisfies

$$(f_0:f)^{-1} \le I_\varphi(f) \le (f:f_0). \tag{10.12}$$

We now show that, in a certain sense, I_φ is approximately additive when $\operatorname{supp}(\varphi)$ is small.

(10.13) Lemma. If $f_1, f_2 \in C_c^+$ and $\varepsilon > 0$, there is a neighborhood V of e such that $I_\varphi(f_1) + I_\varphi(f_2) \le I_\varphi(f_1 + f_2) + \varepsilon$ whenever $\operatorname{supp}(\varphi) \subset V$.

Proof. Fix $g \in C_c^+$ such that $g = 1$ on $\operatorname{supp}(f_1 + f_2)$ and let δ be a positive number. Set $h = f_1 + f_2 + \delta g$ and $h_i = f_i/h$ $(i = 1, 2)$, where it is understood that $h_i = 0$ outside $\operatorname{supp}(f_i)$. Then $h_i \in C_c^+$, so by Proposition (10.2) there is a neighborhood V of e such that $|h_i(x) - h_i(y)| < \delta$ if $i = 1, 2$ and $y^{-1}x \in V$. If $\varphi \in C_c^+$, $\operatorname{supp}(\varphi) \subset V$, and $h \le \Sigma_1^n c_j L_{x_j}\varphi$, then $|h_i(x) - h_i(x_j)| < \delta$ whenever $x_j^{-1}x \in \operatorname{supp}(\varphi)$, so

$$f_i(x) = h(x)h_i(x) \le \sum_j c_j\varphi\left(x_j^{-1}x\right)h_i(x) \le \sum_j c_j\varphi\left(x_j^{-1}x\right)\left[h_i(x_j) + \delta\right].$$

But then $(f_i:\varphi) \le \Sigma_j c_j[h_i(x_j) + \delta]$, and since $h_1 + h_2 \le 1$,

$$(f_1:\varphi) + (f_2:\varphi) \le \sum_j c_j[1 + 2\delta].$$

Now, $\Sigma_j c_j$ can be made arbitrarily close to $(h:\varphi)$, so by (10.7),

$$I_\varphi(f_1) + I_\varphi(f_2) \le (1 + 2\delta)I_\varphi(h) \le (1 + 2\delta)\left[I_\varphi(f_1 + f_2) + \delta I_\varphi(g)\right].$$

In view of (10.12), therefore, if we choose δ small enough so that

$$2\delta(f_1 + f_2 : f_0) + \delta(1 + 2\delta)(g:f_0) < \varepsilon,$$

we are done. □

Proof of Theorem (10.5). For each $f \in C_c^+$ let X_f be the interval $[(f_0:f)^{-1}, (f:f_0)]$, and let $X = \prod_{f \in C_c^+} X_f$. Then X is a compact Hausdorff space by Tychonoff's theorem, and by (10.12), every I_φ is an element of X. For each compact neighborhood V of e, let K_V be the closure in X of $\{I_\varphi : \operatorname{supp}(\varphi) \subset V\}$. Clearly $\bigcap_1^n K_{V_j} \supset K_{\bigcap_1^n V_j}$, so by Proposition (4.21) there is an element I in the intersection of the K_V's. Every neighborhood of I in X intersects $\{I_\varphi : \operatorname{supp}(\varphi) \subset V\}$ for all V. In other words, for any neighborhood V of e and any $f_1, \dots, f_n \in C_c^+$ and $\varepsilon > 0$ there exists $\varphi \in C_c^+$ with $\operatorname{supp}(\varphi) \subset V$ such that $|I(f_j) - I_\varphi(f_j)| < \varepsilon$ for $j = 1, \dots, n$. Therefore, in view of (10.6)–(10.8) and Lemma (10.13), I is left-invariant and satisfies $I(c_1 f_1 + c_2 f_2) = c_1 I(f_1) + c_2 I(f_2)$ for any $c_1, c_2 \ge 0$, and it follows easily [as in the proof of Lemma (7.15)] that if we define $I(f) = I(f^+) - I(f^-)$ for $f \in C_c(G)$, then I is a left-invariant positive linear functional on $C_c(G)$. Moreover, $I(f) > 0$ for all $f \in C_c^+$ by (10.12). The proof is therefore completed by invoking the Riesz representation theorem. □

Having established the existence of Haar measure, we now examine its uniqueness.

(10.14) Theorem. If μ and ν are left Haar measures on G, there exists $c > 0$ such that $\mu = c\nu$.

Proof. We first present a simple proof which works when μ is both left- and right-invariant—in particular, when G is Abelian. Pick $g \in C_c^+$ and set $h(x) = g(x) + g(x^{-1})$, so that $h \in C_c^+$ and $h(x) = h(x^{-1})$. Then for any

$f \in C_c(G)$,

$$\left(\int h\,d\nu\right)\left(\int f\,d\mu\right) = \iint h(y)f(x)\,d\mu(x)\,d\nu(y)$$

$$= \iint h(y)f(xy)\,d\mu(x)\,d\nu(y)$$

$$= \iint h(y)f(xy)\,d\nu(y)\,d\mu(x)$$

$$= \iint h(x^{-1}y)f(y)\,d\nu(y)\,d\mu(x)$$

$$= \iint h(y^{-1}x)f(y)\,d\nu(y)\,d\mu(x)$$

$$= \iint h(y^{-1}x)f(y)\,d\mu(x)\,d\nu(y)$$

$$= \iint h(x)f(y)\,d\mu(x)\,d\nu(y)$$

$$= \left(\int h\,d\mu\right)\left(\int f\,d\nu\right),$$

so that $\mu = c\nu$ where c is the ratio of $\int h\,d\mu$ to $\int h\,d\nu$. [$\int h\,d\nu \neq 0$ by Proposition (10.4c), and Tonelli's theorem is applicable since the functions in question are supported in sets which are compact and hence of finite measure. The same remarks apply below.]

Now, another proof for the general case. The assertion that $\mu = c\nu$ is equivalent to the assertion that $\int f\,d\mu / \int f\,d\nu$ is independent of $f \in C_c^+$. Suppose then that $f, g \in C_c^+$; we shall show that $\int f\,d\mu / \int f\,d\nu = \int g\,d\mu / \int g\,d\nu$. Fix a symmetric compact neighborhood V_0 of e and set

$$A = (\operatorname{supp}(f))V_0 \cup V_0(\operatorname{supp}(f)), \quad B = (\operatorname{supp}(g))V_0 \cup V_0(\operatorname{supp}(g)).$$

Then A and B are compact by Proposition (10.1f), and for $y \in V_0$ the functions $x \to f(xy) - f(yx)$ and $x \to g(xy) - g(yx)$ are supported in A and B, respectively. Next, given $\varepsilon > 0$, by Proposition (10.2) there is a symmetric compact neighborhood $V \subset V_0$ of e such that $\sup_x |f(xy) - f(yx)| < \varepsilon$ and $\sup_x |g(xy) - g(yx)| < \varepsilon$ for $y \in V$. Pick $h \in C_c^+$ with $\operatorname{supp}(h) \subset V$ and $h(x) = h(x^{-1})$. Then

$$\left(\int h\,d\nu\right)\left(\int f\,d\mu\right) = \iint h(y)f(x)\,d\mu(x)\,d\nu(y)$$

$$= \iint h(y)f(yx)\,d\mu(x)\,d\nu(y),$$

and since $h(x) = h(x^{-1})$,

$$\left(\int h\,d\mu\right)\left(\int f\,d\nu\right) = \iint h(x)f(y)\,d\mu(x)\,d\nu(y)$$

$$= \iint h(y^{-1}x)f(y)\,d\mu(x)\,d\nu(y)$$

$$= \iint h(x^{-1}y)f(y)\,d\nu(y)\,d\mu(x)$$

$$= \iint h(y)f(xy)\,d\nu(y)\,d\mu(x)$$

$$= \iint h(y)f(xy)\,d\mu(x)\,d\nu(y).$$

Thus,

$$\left|\left(\int h\,d\mu\right)\left(\int f\,d\nu\right) - \left(\int h\,d\nu\right)\left(\int f\,d\mu\right)\right|$$

$$= \left|\iint h(y)[f(xy) - f(yx)]\,d\mu(x)\,d\nu(y)\right|$$

$$\leq \varepsilon\mu(A)\int h\,d\nu.$$

By the same reasoning,

$$\left|\left(\int h\,d\mu\right)\left(\int g\,d\nu\right) - \left(\int h\,d\nu\right)\left(\int g\,d\mu\right)\right| \leq \varepsilon\mu(B)\int h\,d\nu.$$

Dividing these inequalities by $(\int h\,d\nu)(\int f\,d\nu)$ and $(\int h\,d\nu)(\int g\,d\nu)$, respectively, and adding, we obtain

$$\left|\frac{\int f\,d\mu}{\int f\,d\nu} - \frac{\int g\,d\mu}{\int g\,d\nu}\right| \leq \varepsilon\left(\frac{\mu(A)}{\int f\,d\nu} + \frac{\mu(B)}{\int g\,d\nu}\right).$$

Since ε is arbitrary, we are done. □

We conclude by investigating the relationship between left and right Haar measure. If μ is a left Haar measure on G and $x \in G$, the measure $\mu_x(E) = \mu(Ex)$ is again a left Haar measure, because of the commutativity of left and right translations (i.e., the associative law). Hence there is a positive number $\Delta(x)$ such that $\mu_x = \Delta(x)\mu$. The function Δ: $G \to (0, \infty)$ thus defined is

independent of the choice of μ by Theorem (10.14); it is called the **modular function** of G. We clearly have

$$\mu_{xy} = (\mu_x)_y \quad \text{and} \quad \int R_x f\,d\mu = \int f\,d\mu_{x^{-1}} = \Delta(x^{-1})\int f\,d\mu \text{ for } f \in C_c(G).$$

It follows that $\Delta(xy) = \Delta(x)\Delta(y)$ and that Δ is continuous (see Exercise 2). In short, Δ is a continuous homomorphism from G to the multiplicative group $(0, \infty)$.

Evidently, a left Haar measure on G is also a right Haar measure precisely when Δ is identically 1, in which case G is called **unimodular**. Of course every Abelian group is unimodular; remarkably enough, groups which are highly noncommutative are also unimodular. More precisely, let $[G, G]$ denote the smallest closed subgroup of G containing all elements of the form $[x, y] = xyx^{-1}y^{-1}$. $[G, G]$ is called the **commutator subgroup** of G; it is normal because $z[x, y]z^{-1} = [zxz^{-1}, zyz^{-1}]$, and it is trivial precisely when G is Abelian.

(10.15) Proposition. If $G/[G, G]$ is finite, then G is unimodular.

Proof. Every continuous homomorphism (such as Δ) from G to an Abelian group must annihilate $[x, y]$ for all x, y and must therefore factor through $G/[G, G]$. If the latter group is finite, $\Delta(G)$ is a finite subgroup of $(0, \infty)$. But $(0, \infty)$ has no finite subgroups except $\{1\}$. □

(10.16) Proposition. If G is compact, then G is unimodular.

Proof. For any $x \in G$, obviously $G = Gx$. Hence if μ is a left Haar measure, $\mu(G) = \mu(Gx) = \Delta(x)\mu(G)$, and since $0 < \mu(G) < \infty$ we conclude that $\Delta(x) = 1$. □

We observed above that if μ is a left Haar measure, $\tilde{\mu}(E) = \mu(E^{-1})$ is a right Haar measure. We now show how to compute $\tilde{\mu}$ in terms of μ and Δ.

(10.17) Proposition. $d\tilde{\mu}(x) = \Delta(x^{-1})\,d\mu(x)$.

Proof. Let $\delta(x) = \Delta(x^{-1})$. If $f \in C_c(G)$,

$$\int f(xy^{-1})\delta(x)\,d\mu(x) = \delta(y)\int f(xy^{-1})\delta(xy^{-1})\,d\mu(x)$$
$$= \delta(y)\int R_y(f\delta)\,d\mu = \int f\delta\,d\mu$$

Hence the functional $f \to \int f\delta\,d\mu$ is right-invariant, so there is a constant c such that $\int f\delta\,d\mu = c\int f\,d\tilde{\mu}$. Given $\varepsilon > 0$, let V be a symmetric neighborhood of e such that $|\delta(x) - 1| < \varepsilon$ for $x \in V$, and choose $f \in C_c^+$ such that $f(x) = f(x^{-1})$ and $\operatorname{supp}(f) \subset V$. Then

$$\int f\,d\tilde{\mu} = \int f\,d\mu \quad \text{and} \quad \left|\int f\delta\,d\mu - \int f\,d\mu\right| \le \varepsilon\int f\,d\mu.$$

It follows that $c = 1$, so the measures $\delta\,d\mu$ and $d\tilde{\mu}$ define the same functional on $C_c(G)$. Bug $\delta\,d\mu$ is a Radon measure (Exercise 7.9), so $d\tilde{\mu} = \delta\,d\mu$ by the uniqueness of the Riesz representation. □

(10.18) Corollary. Left and right Haar measures are mutually absolutely continuous.

EXERCISES

1. If G is a topological group and $E \subset G$, then $\bar{E} = \bigcap\{EV: V \text{ is a neighborhood of } e\}$.
2. If μ is a Radon measure on the locally compact group G and $f \in C_c(G)$, the functions $x \to \int L_x f\,d\mu$ and $x \to \int R_x f\,d\mu$ are continuous.
3. Let G be a locally compact group which is homeomorphic to an open subset U of $\mathbf{R}^n$ in such a way that, if we identify G with U, left translation is an affine map—that is, $xy = A_x(y) + b_x$ where A_x is a linear transformation of $\mathbf{R}^n$ and $b_x \in \mathbf{R}^n$. Then $|\det A_x|^{-1}\,dx$ is a left Haar measure on G, where dx denotes Lebesgue measure on $\mathbf{R}^n$. (Similarly for right translations and right Haar measures.)
4. The following are special cases of Exercise 3.
   (a) If G is the multiplicative group $(0, \infty)$, $x^{-1}\,dx$ is a Haar measure.
   (b) If G is the group of invertible $n \times n$ real matrices, $|\det A|^{-n}\,dA$ is a left and right Haar measure, where dA = Lebesgue measure on $\mathbf{R}^{n^2}$. [To see that the determinant of the map $X \to AX$ on matrix space is $|\det A|^n$, consider the case where A is an elementary matrix as in the proof of Theorem (2.43).]
   (c) If G is the group of 3×3 matrices of the form

$$\begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} \quad (x, y, z \in \mathbf{R}),$$

   then $dx\,dy\,dz$ is a left and right Haar measure.
   (d) If G is the group of 2×2 matrices of the form

$$\begin{pmatrix} x & y \\ 0 & 1 \end{pmatrix} \quad (x > 0, y \in \mathbf{R}),$$

   then $x^{-2}\,dx\,dy$ is a left Haar measure and $x^{-1}\,dx\,dy$ is a right Haar measure.
5. Let G be as in Exercise 4d. Construct a Borel set in G with finite left Haar measure but infinite right Haar measure, and a left uniformly continuous function on G that is not right uniformly continuous.

6. Let $\{G_\alpha\}_{\alpha\in A}$ be a family of topological groups and $G = \prod_{\alpha\in A} G_\alpha$.
   (a) With the product topology and coordinatewise multiplication, G is a topological group.
   (b) If each G_α is compact and μ_α is the Haar measure on G_α such that $\mu_\alpha(G_\alpha) = 1$, then the Radon product of the μ_α's, as constructed in Theorem (9.19), is a Haar measure on G.

7. In Exercise 6, let G_α be the multiplicative group $\{-1, 1\}$ with the discrete topology, for all α. Let μ be a Haar measure on G.
   (a) If $\pi_\alpha: G \to \{-1, 1\}$ is the αth coordinate function, then $\int \pi_\alpha \pi_\beta \, d\mu = 0$ for $\alpha \neq \beta$.
   (b) If A is uncountable, $L^2(\mu)$ is not separable, even though $\mu(G) < \infty$.

8. $\mathbf{Q}$ is a topological group with the relative topology induced from $\mathbf{R}$. It is not locally compact, and there is no nonzero translation-invariant Borel measure on $\mathbf{Q}$ which is finite on compact sets.

9. Let G be a locally compact group with left Haar measure μ.
   (a) G has a subgroup H which is open, closed, and σ-compact. (Let H be the subgroup generated by a precompact open neighborhood of e.)
   (b) The restriction of μ to subsets of H is a left Haar measure on H.
   (c) μ is decomposable in the sense of Exercise 3.15.
   (d) μ is regular iff μ is semifinite iff μ is σ-finite iff G is σ-compact. (Cf. Exercises 7.13 and 7.14.)

2. HAUSDORFF MEASURE

In geometric problems it is important to have a method for measuring the size of lower-dimensional sets in $\mathbf{R}^n$, such as curves and surfaces in $\mathbf{R}^3$. Differential-geometric techniques provide such a method which applies to smooth submanifolds of $\mathbf{R}^n$ (see Section 3). However, there is also a measure-theoretic approach to the problem which yields more general results. Indeed, the basic ideas can be carried out just as easily in arbitrary metric spaces, so we begin by working in this generality.

Let (X, ρ) be a metric space. (See the Prologue for the relevant terminology.) An outer measure μ^* on x is called a **metric outer measure** if

$$\mu^*(A \cup B) = \mu^*(A) + \mu^*(B) \text{ whenever } \rho(A, B) > 0.$$

(10.19) Proposition. If μ^* is a metric outer measure on X, then every closed subset of X, and hence every Borel subset of X, is μ^*-measurable.

Proof. We must show that if F is closed and $A \subset X$ satisfies $\mu^*(A) < \infty$, then $\mu^*(A) \geq \mu^*(A \cap F) + \mu^*(A \setminus F)$. Let $B_n = \{x \in A \setminus F: \ \rho(x, F) \geq$

$n^{-1}\}$. Then $\{B_n\}$ is an increasing sequence of sets whose union is $A \setminus F$ (since F is closed), and $\rho(B_n, F) \geq n^{-1}$. Therefore,

$$\mu^*(A) \geq \mu^*((A \cap F) \cup B_n) = \mu^*(A \cap F) + \mu^*(B_n),$$

so it will suffice to show that $\mu^*(A \setminus F) = \lim \mu^*(B_n)$. Let $C_n = B_{n+1} \setminus B_n$. If $x \in C_{n+1}$ and $\rho(x, y) < [n(n+1)]^{-1}$, then

$$\rho(y, F) \leq \rho(y, x) + \rho(x, F) < \frac{1}{n(n+1)} + \frac{1}{n+1} = \frac{1}{n},$$

so that $\rho(C_{n+1}, B_n) \geq [n(n+1)]^{-1}$. A simple induction therefore shows that

$$\mu^*(B_{2k+1}) \geq \mu^*(C_{2k} \cup B_{2k-1}) = \mu^*(C_{2k}) + \mu^*(B_{2k-1})$$

$$\geq \cdots \geq \sum_1^k \mu^*(C_{2j}),$$

and similarly $\mu^*(B_{2k}) \geq \sum_1^k \mu^*(C_{2j-1})$. Since $\mu^*(B_n) \leq \mu^*(A) < \infty$ it follows that the series $\sum_1^\infty \mu^*(C_{2j})$ and $\sum_1^\infty \mu^*(C_{2j-1})$ are convergent. But by subadditivity we have

$$\mu^*(A \setminus F) \leq \mu^*(B_n) + \sum_{n+1}^\infty \mu^*(C_j),$$

so as $n \to \infty$ the last sum vanishes and we obtain

$$\mu^*(A \setminus F) \leq \liminf \mu^*(B_n) \leq \limsup \mu^*(B_n) \leq \mu^*(A \setminus F),$$

as desired. □

We are now ready to define Hausdorff measure. Suppose (X, ρ) is a metric space, $p \geq 0$, and $\delta > 0$. If $A \subset X$, let

$$H_{p,\delta}(A) = \inf\left\{\sum_1^\infty (\operatorname{diam} B_j)^p: \ A \subset \cup_1^\infty B_j \text{ and } \operatorname{diam} B_j \leq \delta\right\},$$

with the convention that $\inf \varnothing = \infty$. As δ decreases the infimum is being taken over a smaller family of coverings of A, so $H_{p,\delta}(A)$ increases. The quantity

$$H_p(A) = \lim_{\delta \to 0} H_{p,\delta}(A)$$

is called the ***p*-dimensional Hausdorff (outer) measure** of A.

Several comments on this definition are in order.

(i) The sets B_j in the definition of $H_{p,\delta}$ are arbitrary subsets of X. However, one obtains the same result if one requires the B_j's to be closed (since $\operatorname{diam} B_j = \operatorname{diam} \overline{B}_j$) or if one requires the B_j's to be open (because any B is contained in an open set U with $\operatorname{diam} U - \operatorname{diam} B$ arbitrarily small). Similarly, if $X = \mathbf{R}$, one can restrict the B_j's to be closed or open intervals.

(ii) The intuition behind the definition of H_p is that if p is an integer and B is a "nice" subset of $\mathbf{R}^p$ like a ball or a cube, the measure of B is roughly $(\operatorname{diam} B)^p$.

(iii) The restriction to coverings by sets of small diameter is necessary to provide an accurate measure of irregularly shaped sets; otherwise one could simply cover a set by itself, with the result that its measure would be at most the pth power of its diameter. Consider, for example, the curve $A_n = \{(x, \sin nx): -1 \le x \le 1\}$ in $\mathbf{R}^2$. Clearly $\operatorname{diam} A_n \le 2^{3/2}$ for all n, but the length of A_n tends to ∞ along with n. One needs to take $\delta \ll n^{-1}$ before $H_{1,\delta}(A)$ becomes an accurate estimate of the length of A_n.

We now derive the basic properties of H_p.

(10.20) Proposition. *H_p is a metric outer measure.*

Proof. $H_{p,\delta}$ is an outer measure by Proposition (1.10), and by letting $\delta \to 0$ one concludes that H_p is an outer measure. If $\rho(A, B) > 0$ and $\{C_n\}$ is a covering of $A \cup B$ such that $(\operatorname{diam} C_n) \le \delta < \rho(A, B)$ for all n, then no C_n can intersect both A and B. Splitting $\sum(\operatorname{diam} C_n)^p$ into two parts according to whether $C_n \cap A = \varnothing$ or $C_n \cap B = \varnothing$ shows that $\sum(\operatorname{diam} C_n)^p \ge H_{p,\delta}(A) + H_{p,\delta}(B)$, and hence $H_{p,\delta}(A \cup B) \ge H_{p,\delta}(A) + H_{p,\delta}(B)$. As this inequality is valid whenever $\delta < \rho(A, B)$, the desired result follows by letting $\delta \to 0$. □

In view of Propositions (10.19) and (10.20), the restriction of H_p to the Borel sets is a measure, which we still denote by H_p and call p-dimensional Hausdorff measure.

(10.21) Proposition. *H_p is invariant under isometries of X. Moreover, if Y is any set and f, g: $Y \to X$ satisfy $\rho(f(y), f(z)) \le C\rho(g(y), g(z))$ for all $y, z \in Y$, then $H_p(f(A)) \le C^p H_p(g(A))$ for all $A \subset Y$.*

Proof. The first statement is obvious from the definition of H_p. As for the second, given $\varepsilon, \delta > 0$, cover $g(A)$ by sets B_j such that $\operatorname{diam} B_j \le C^{-1}\delta$ and $\sum(\operatorname{diam} B_j)^p \le H_p(g(A)) + \varepsilon$. The the sets $B_j' = f(g^{-1}(B_j))$ cover $f(A)$, and $\operatorname{diam} B_j' \le C(\operatorname{diam} B_j) \le \delta$, so that

$$H_{p,\delta}(f(A)) \le \sum \left(\operatorname{diam} B_j'\right)^p \le C^p H_p(g(A)) + C^p\varepsilon.$$

The proof is completed by letting $\delta, \varepsilon \to 0$. □

(10.22) Proposition. *If $H_p(A) < \infty$, then $H_p(A) = 0$ for all $q > p$. If $H_p(A) > 0$, then $H_q(A) = \infty$ for all $q < p$.*

Proof. It suffices to prove the first statement, as the second one is its contrapositive. If $H_p(A) < \infty$, for any $\delta > 0$ there exist B_j with $A \subset \bigcup B_j$, $\operatorname{diam} B_j \le \delta$, and $\sum(\operatorname{diam} B_j)^p \le H_p(A) + 1$. But if $q > p$,

$$\sum (\operatorname{diam} B_j)^q \le \delta^{q-p} \sum (\operatorname{diam} B_j)^p \le \delta^{q-p}\left[H_p(A) + 1\right],$$

so $H_{q,\delta}(A) \le \delta^{q-p}[H_p(A) + 1]$. Letting $\delta \to 0$, we see that $H_q(A) = 0$. □

According to Proposition (10.22), for any $A \subset X$ the numbers

$$\inf\{p \ge 0: H_p(A) = 0\}, \qquad \sup\{p \ge 0: H_p(A) = \infty\}$$

are equal. Their common value is called the **Hausdorff dimension** of A.

From now on we restrict attention to the case $X = \mathbf{R}^n$. Our object is to show that if $p = 1, \ldots, n$, then H_p gives the geometrically correct notion of measure (modulo a normalization constant) for p-dimensional submanifolds of $\mathbf{R}^n$. We begin with the case $p = n$.

(10.23) Proposition. *There is a constant $\gamma_n > 0$ such that $\gamma_n H_n$ is Lebesgue measure on $\mathbf{R}^n$.*

Proof. H_n is a translation-invariant Borel measure on $\mathbf{R}^n$. If $Q \subset \mathbf{R}^n$ is a cube, it is easily verified that $0 < H_n(Q) < \infty$ (Exercise 10). It follows that $H_n \ne 0$ and that H_n is finite on compact sets, whence H_n is a Radon measure by Theorem (7.8). The desired result is therefore a consequence of Theorem (10.14). (The simple argument given there for the Abelian case can be read without going through the rest of Section 1: simply read $x + y$ for xy and $-x$ for x^{-1}.) □

The normalization constant γ_n turns out to be the volume of a ball of diameter 1, which by Corollary (2.56) is $\pi^{n/2}/2^n\Gamma((n/2) + 1)$. We shall not give the proof here, as it is rather complicated, and the value of γ_n is irrelevant for our purposes. Many authors build the constant γ_n into the definition of Hausdorff measure; that is, they define p-dimensional Hausdorff measure, for $p \in [0, \infty)$, to be $\gamma_p H_p$, where $\gamma_p = \pi^{p/2}/2^p\Gamma((p/2) + 1)$.

We now consider lower-dimensional sets in $\mathbf{R}^n$. If $1 \le k \le n$, a ***k*-dimensional C^1 submanifold** of $\mathbf{R}^n$ is a set $M \subset \mathbf{R}^n$ with the following property: for each $x \in M$ there exist a neighborhood U of x in $\mathbf{R}^n$, an open set $V \subset \mathbf{R}^k$, and an injective map f: $V \to U$ of class C^1 such that $f(V) = M \cap U$ and the differential $D_x f$ (i.e., the linear map from $\mathbf{R}^k$ to $\mathbf{R}^n$ whose matrix is $[(\partial f_i/\partial x_j)(x)]$) is injective for each $x \in V$. Such an f is called a **parameterization** of $M \cap U$. Every submanifold M can be covered by countably many U's

for which $M \cap U$ has a parametrization, so for our purposes it will suffice to assume that $M = M \cap U$ has a global parametrization.

(There are other common definitions of "submanifold": M is a k-dimensional C^1 submanifold if it is locally the set of zeros of a C^1 map g: $\mathbf{R}^n \to \mathbf{R}^{n-k}$ such that $D_x g$ is surjective at each $x \in M$, or if it is locally the image of $\{x \in \mathbf{R}^n\colon x_j = 0 \text{ for } j > k\}$ under a C^1 diffeomorphism of $\mathbf{R}^n$. The equivalence of these definitions with the one given above is a standard exercise in the use of the implicit function theorem.)

We begin our study of submanifolds with the linear case. If T is a linear map from $\mathbf{R}^k$ to $\mathbf{R}^n$ and $T^*\colon \mathbf{R}^n \to \mathbf{R}^k$ is its adjoint, then T^*T is a positive semidefinite operator on $\mathbf{R}^k$. Its determinant is therefore nonnegative, and we may define

$$J(T) = \sqrt{\det(T^*T)}\,.$$

(10.24) Proposition. If $k \le n$, $T\colon \mathbf{R}^k \to \mathbf{R}^n$ is linear, and $A \subset \mathbf{R}^k$, then $H_k(T(A)) = J(T)H_k(A)$.

Proof. If $k = n$, then $\det(T^*T) = (\det T)^2$, so $J(T) = |\det T|$ and the assertion reduces to Theorem (2.43) because of Proposition (10.23). If $k < n$, let R be a rotation of $\mathbf{R}^n$ which maps the range of T to the subspace $\mathbf{R}^k \times \{0\} = \{y \in \mathbf{R}^n\colon y_j = 0 \text{ for } j > k\}$, and let $S = RT$. Then $S^*S = T^*R^*RT = T^*T$, so that $J(S) = J(T)$, and $H_k(S(A)) = H_k(T(A))$ since H_k is rotation-invariant. But if we identify $\mathbf{R}^k \times \{0\}$ with $\mathbf{R}^k$, S becomes a map from $\mathbf{R}^k$ to itself, and the definition of S^*S is unchanged when this identification is made. We are therefore back in the equidimensional case, which was disposed of above. □

It is now easy to guess what the corresponding formula must be for a general smooth injection $f\colon \mathbf{R}^k \to \mathbf{R}^n$, since locally every smooth map is approximately linear. Our next lemma makes this idea precise.

(10.25) Lemma. Suppose M is a k-dimensional C^1 submanifold of $\mathbf{R}^n$ parametrized by $f\colon V \to \mathbf{R}^n$. For any $\alpha > 1$ there is a sequence $\{B_j\}$ of disjoint Borel subsets of V such that $V = \bigcup_1^\infty B_j$, and a sequence $\{T_j\}$ of linear maps from $\mathbf{R}^k$ to $\mathbf{R}^n$, such that

$$\alpha^{-1}|T_j z| \le |D_x f(z)| \le \alpha |T_j z| \quad \text{for } x \in B_j,\ z \in \mathbf{R}^k; \tag{10.26}$$

$$\alpha^{-1}|T_j x - T_j y| \le |f(x) - f(y)| \le \alpha |T_j x - T_j y| \quad \text{for } x, y \in B_j. \tag{10.27}$$

Proof. Let us fix $\varepsilon > 0$ and $\beta > 1$ such that

$$\alpha^{-1} + \varepsilon < \beta^{-1} < 1 < \beta < \alpha - \varepsilon,$$

and let $\mathscr{T}$ be a countable dense subset of the set of linear maps from $\mathbf{R}^k$ to $\mathbf{R}^n$

(e.g., the set of matrices with rational entries). For $T \in \mathscr{T}$ and $m \in \mathbf{N}$ let $E(T,m)$ be the set of all $x \in V$ such that

$$\beta^{-1}|Tz| \le |D_x f(z)| \le \beta |Tz| \quad \text{for } z \in \mathbf{R}^k,$$

$$\alpha^{-1}|Tx - Ty| \le |f(x) - f(y)| \le \alpha |Tx - Ty|$$

$$\text{for all } y \in V \text{ such that } |y - x| < m^{-1}.$$

$E(T,m)$ is a Borel set since one can restrict y, z to lie in countable dense sets. It will therefore suffice to show that the sets $E(T,m)$ cover V. Indeed, each $E(T,m)$ is a countable union of sets $E_i(T,m)$ of diameter less than m^{-1}, so by disjointifying the countable collection $\{E_i(T,m)\colon T \in \mathscr{T},\ i, m \in \mathbf{N}\}$ we obtain the desired sets B_j and the associated linear maps T_j.

Suppose then that $x \in V$. Since $D_x f$ is injective, the number $\delta_0 = \inf\{|D_x f(z)|\colon |z| = 1\}$ is positive. Choose $\delta > 0$ so that $\delta \le (\beta - 1)\delta_0$ and $\delta \le (1 - \beta^{-1})\delta_0$, and then pick $T \in \mathscr{T}$ such that $\|T - D_x f\| < \delta$. Then

$$|Tz| \le |D_x f(z)| + |Tz - D_x f(z)| \le |D_x f(z)| + \delta|z| \le \beta |D_x f(z)|,$$

and similarly $|Tz| \ge \beta^{-1}|D_x f(z)|$. This establishes the first inequality and also shows that T is injective, so $\eta = \inf\{|Tz|\colon |z| = 1\}$ is positive. Since f is differentiable at x, there exists $m \in \mathbf{N}$ such that

$$|f(y) - f(x) - D_y f(y - x)| \le \varepsilon\eta |y - x| \le \varepsilon |T(y - x)| \quad \text{for } |x - y| < m^{-1}.$$

But then

$$|f(y) - f(x)| \le |f(y) - f(x) - D_x f(y - x)| + |D_x f(y - x)|$$

$$\le \varepsilon |Ty - Tx| + \beta |Ty - Tx| < \alpha |Ty - Tx|,$$

and similarly $|f(y) - f(x)| \ge \alpha^{-1}|Ty - Tx|$. In short, $x \in E(T,m)$, so we are done. □

(10.28) Theorem. Suppose M is a k-dimensional C^1 submanifold of $\mathbf{R}^n$ parametrized by $f\colon V \to \mathbf{R}^n$. If A is a Borel subset of V, then $f(A)$ is a Borel subset of $\mathbf{R}^n$, and

$$H_k(f(A)) = \int_A J(D_x f)\, dH_k(x). \tag{10.29}$$

Moreover, If φ is a Borel measurable function on M which is nonnegative or in $L^1(M, H_k)$, then

$$\int_M \varphi(y)\, dH_k(y) = \int_V \varphi(f(x))\, J(D_x f)\, dH_k(x). \tag{10.30}$$

Proof. Since V is an open subset of $\mathbf{R}^k$, it is σ-compact. It follows that if A is closed in V, then A is σ-compact, and hence $f(A)$ is σ-compact since f is continuous. The collection of all $A \subset V$ such that $f(A)$ is Borel is therefore a σ-algebra which includes all closed sets, hence all Borel sets. We shall prove (10.29); this establishes (10.30) when $\varphi = \chi_{f(A)}$, and the general result follows from the usual linearity and approximation arguments.

Given $\alpha > 1$, let $\{B_j\}$, $\{T_j\}$ be as in Lemma (10.25), and let $A_j = A \cap B_j$. It follows from (10.26) and Proposition (10.21) that

$$\alpha^{-k} H_k\big(T_j(E)\big) \le H_k\big(D_x f(E)\big) \le \alpha^k H_k\big(T_j(E)\big) \quad \text{for } x \in A_j,\ E \subset \mathbf{R}^k,$$

and hence by Proposition (10.24),

$$\alpha^{-k} J(T_j) \le J(D_x f) \le \alpha^k J(T_j) \quad \text{for } x \in A_j.$$

But it also follows from (10.27) and Proposition (10.21) that

$$\alpha^{-k} H_k\big(T_j(A_j)\big) \le H_k\big(f(A_j)\big) \le \alpha^k H_k\big(T_j(A_j)\big)$$

and from Proposition (10.24) that $H_k(T_j(A_j)) = J(T_j) H_k(A_j)$. Therefore,

$$\alpha^{-2k} H_k\big(f(A_j)\big) \le \alpha^{-k} J(T_j) H_k(A_j) \le \int_{A_j} J(D_x f)\, dH_k(x)$$

$$\le \alpha^k J(T_j) H_k(A_j) \le \alpha^{2k} H_k\big(f(A_j)\big).$$

Summing over j, we obtain

$$\alpha^{-2k} H_k(f(A)) \le \int_A J(D_x f)\, dH_k(x) \le \alpha^{2k} H_k(f(A)),$$

so the proof is completed by letting $\alpha \to 1$. □

If both sides of the equations (10.29) and (10.30) are multiplied by the normalizing constant γ_k in Proposition (10.23), the integrals on the right become ordinary Lebesgue integrals, and we obtain the formulas for measures and integrals on M given by Riemannian geometry (see Section 3). Moreover, if $k = n$ we have $J(D_x f) = |\det D_x f|$, so the result reduces to Theorem (2.47). See also Exercises 11 and 12.

We now show that the theory of p-dimensional Hausdorff measure for $p \notin \mathbf{N}$ is nonvacuous by displaying, for any $p \in (0,1)$, a set $B = B_p \subset [0,1]$ such that $H_p(B) = 1$. B will be a Cantor-type set, constructed as follows. Fix $p \in (0,1)$ and set $\xi_n = 2^{-n/p}$; the numbers $\xi_n - 2\xi_{n+1}$ are thus positive for all n. Remove from $[0,1]$ the open middle interval of length $1 - 2\xi_1$, denoted by

$I(\frac{1}{2})$, obtaining a set B_1 which is the union of the two closed intervals $[0, \xi_1]$ and $[1 - \xi_1, 1]$. Remove from each of these intervals the open middle intervals of length $\xi_1 - 2\xi_2$, denoted by $I(\frac{1}{4})$ and $I(\frac{3}{4})$, obtaining a set B_2 which is the union of four closed intervals of length ξ_2. Now proceed inductively: having constructed B_{n-1}, remove from its 2^{n-1} constituent intervals the open middle intervals of length $\xi_{n-1} - 2\xi_n$, denoted by $I(2^{-n}), I(3 \cdot 2^{-n}), \ldots, I((2n-1)2^{-n})$, to obtain B_n. Let $B = \bigcap_1^\infty B_n$. Then B is a generalized Cantor set, and if $p = \log_3 2$ so that $\xi_n = 3^{-n}$, B is the ordinary Cantor set.

(10.31) Theorem. With notation as above, $H_p(B) = 1$.

Proof. The closed intervals $B_{n1}, \ldots, B_{n2^n}$ which make up B_n cover B, and $\operatorname{diam}(B_{nj}) = \xi_n \to 0$ as $n \to \infty$. Since

$$\sum_1^{2^n} \big(\operatorname{diam} B_{nj}\big)^p = 2^n \xi_n^p = 1,$$

we have $H_p(B) \le 1$. The reverse inequality is much more difficult to prove; we begin with the following lemma.

(10.32) Lemma. If $x_1 > x_2 > x_3 > 0$, then $(x_1 - x_2 + x_3)^p > x_1^p - x_2^p + x_3^p$.

Proof. Since $p < 1$, the function $f(x) = x^p$ is strictly concave, which means that for $a \le b$ the average slope $[f(b) - f(a)]/(b - a)$ of f on $[a, b]$ is a strictly decreasing function of b (for fixed a) and of a (for fixed b). Clearly $x_1 > x_1 - x_2 + x_3 > x_3$, so

$$\frac{(x_1 - x_2 + x_3)^p - x_3^p}{x_1 - x_2} > \frac{x_1^p - x_3^p}{x_1 - x_3} > \frac{x_1^p - x_2^p}{x_1 - x_2}.$$

The result follows upon multiplying through by $x_1 - x_2$. □

For each dyadic rational $y \in (0,1)$ (i.e., $y = k2^{-n}$ for some $k, n \in \mathbf{N}$), let $u(y)$, $v(y)$ be the left and right endpoints of $I(y)$, and set $v(0) = 0$, $u(1) = 1$. Then since an interval of length ξ_n lies between $I((k-1)2^{-n})$ and $I(k2^{-n})$, we have

$$(10.33) \quad u(k2^{-n}) - v((k-1)2^{-n}) = \xi_n \quad \text{for } 1 \le k \le 2^n,\ n \in \mathbf{N}.$$

A bit of terminology: if $y = k2^{-n}$ where k is odd, we call 2^n the **denominator** of y.

(10.34) Lemma. If y and z are dyadic rationals in $[0,1]$ with $y < z$, and $\alpha = u(z) - v(y)$, then $\alpha^p \ge z - y$.

Proof. We prove this by induction on the denominators of y and z. For $y = 0$, $z = 1$ we have $\alpha^p = 1 = z - y$, and for $y = 0$, $z = \frac{1}{2}$ or $y = \frac{1}{2}$, $z = 1$ we have $\alpha = \xi_1$ and $\alpha^p = \frac{1}{2} = z - y$. We therefore assume the result to be true when the denominators of y and z are at most 2^n, and we consider the situation when the denominators of y and z are at most 2^{n+1} with equality for at least one of them.

Case I: the denominator of y is 2^{n+1}, the denominator of z is $\le 2^n$. Let $y_1 = y - 2^{-n-1}$ and $y_2 = y + 2^{-n-1}$ be the nearest neighbors of y with denominator $\le 2^n$: thus $y_1 < y < y_2 \le z$. By (10.33), $u(y_2) - v(y_1) = \xi_n$ and $u(y_2) - v(y) = \xi_{n+1}$, so $v(y) = v(y_1) + \xi_n - \xi_{n+1}$. If $\alpha = u(z) - v(y)$, then, we have $\alpha = \beta - \xi_n + \xi_{n+1}$ where

$$\beta = u(z) - v(y_1) \ge u(y_2) - v(y_1) = \xi_n.$$

By Lemma (10.32) and the inductive hypothesis,

$$\alpha^p \ge \beta^p - \xi_n^p + \xi_{n+1}^p \ge z - y_1 - 2^{-n} + 2^{-n-1} = z - y.$$

Case II: the denominator of y is $\le 2^n$, the denominator of z is 2^{n+1}. Arguing exactly as in Case I, we set $z_1 = z - 2^{-n-1}$, $z_2 = z + 2^{-n-1}$ and deduce that $\alpha = \beta - \xi_n + \xi_{n+1}$ where $\beta = u(z_2) - v(y) \ge \xi_n$, so that $\alpha^p \ge \beta^p - \xi_n^p + \xi_{n+1}^p \ge z - y$.

Case III: the denominators of y and z are both 2^{n+1}. Let y_1, y_2, z_1, z_2 be as in Cases I and II, so that $y_1 < y_2 \le z_1 < z_2$, $v(y) = v(y_1) + \xi_n - \xi_{n+1}$, and $u(z) = u(z_2) - \xi_n + \xi_{n+1}$. Then

$$\alpha = u(z) - v(y) = \beta - 2\xi_n + 2\xi_{n+1}, \text{ where}$$

$$\beta = u(z_2) - v(y_1) > u(z_2) - v(z_1) + u(y_2) - v(y_1) = 2\xi_n.$$

Hence by Lemma (10.32) and the inductive hypothesis,

$$\alpha^p > \beta^p - 2^p(\xi_n^p - \xi_{n+1}^p) > \beta^p - 2(\xi_n^p - \xi_{n+1}^p)$$

$$\ge z_2 - y_1 - 2^{-n+1} + 2^{-n} = z - y. \qquad \square$$

We can now complete the proof of Theorem (10.31). Given $\delta, \varepsilon > 0$, there is a collection $\{U_j\}_1^J$ of open intervals with lengths $\{\eta_j\}_1^J$ ($\eta_j < \delta$) which cover B such that $\sum_1^J \eta_j^p < H_{p,\delta}(B) + \varepsilon$; the collection may be assumed finite since B is compact. We proceed to modify the intervals U_j as follows. First, we replace each U_j by $V_j = \overline{U}_j \cap [0,1]$. Second, if an endpoint of V_j lies in an interval $I(y)$, we replace V_j by $W_j = V_j \setminus I(y)$ (which is still a closed interval). Third, if two or more of the W_j's intersect in pairs we replace them by their union, which is again a closed interval. The result of this is to produce a covering

$\{A_n\}_0^N$ of B by disjoint closed intervals, which we may assume ordered from left to right, with the following properties. First, if α_n is the length of A_n, then

$$\sum_0^N \alpha_n^p \le \sum_1^J \eta_j^p < H_{p,\delta}(B) + \varepsilon.$$

Indeed, the first two operations on the U_j's decrease their lengths, while the third decreases the sum of the pth powers of their lengths because $(a+b)^p < a^p + b^p$. Second, $[0,1] \setminus \bigcup_0^N A_n$ is a finite disjoint union of intervals $I(y_1), \dots, I(y_N)$ with $y_1 < \cdots < y_N$, so that the left (resp. right) endpoint of $I(y_n)$ is the right (resp. left) endpoint of A_{n-1} (resp. A_n). By Lemma (10.34), therefore, we have $\alpha_0^p \ge y_1$, $\alpha_n^p \ge y_n - y_{n-1}$ for $0 < n < N$, and $\alpha_N^p \ge 1 - y_N$. Adding these inequalities, we obtain

$$1 \le \sum_0^N \alpha_n^p < H_{p,\delta}(B) + \varepsilon.$$

Thus $H_{p,\delta}(B) \ge 1$ for all δ, and therefore $H_p(B) = 1$. $\square$

EXERCISES

10. If Q is a cube in $\mathbf{R}^n$, then $0 < H_n(Q) < \infty$. [To show $H_n(Q) < \infty$, consider coverings of Q obtained by dividing Q into congruent subcubes. To show that $H_n(Q) > 0$, observe that any set of diameter d is contained in a ball of diameter $2d$, which has Lebesgue measure proportional to d^n.]

11. If $f\colon (a,b) \to \mathbf{R}^n$ is a parametrization of a smooth curve (i.e., a one-dimensional C^1 submanifold of $\mathbf{R}^n$), the Hausdorff one-dimensional measure of the curve is $\int_a^b |f'(t)|\,dt$.

12. (a) Given $a \in \mathbf{R}^k$, define $T\colon \mathbf{R}^k \to \mathbf{R}^{k+1}$ by $Tx = (x, a \cdot x)$. Then $T^*T = I + S$ where $Sx = (a \cdot x)a$.

(b) $\det(T^*T) = 1 + |a|^2$. (The determinant of T^*T is the product of its eigenvalues. What are the eigenvalues of S?)

(c) If $\varphi\colon \mathbf{R}^k \to \mathbf{R}$ is a C^1 function, the graph of φ is a k-dimensional C^1 submanifold of $\mathbf{R}^{k+1}$ parametrized by $f(x) = (x, \varphi(x))$. If $A \subset \mathbf{R}^k$, the k-dimensional volume of the portion of the graph lying above A is

$$\int_A \sqrt{1 + |\operatorname{grad} \varphi(x)|^2}\,dx.$$

13. In any metric space, zero-dimensional Hausdorff measure is counting measure.

14. If A_m is a subset of a metric space X of Hausdorff dimension p_m for $m \in \mathbf{N}$, then $\bigcup_1^\infty A_m$ has Hausdorff dimension $\sup\{p_m\}$.

15. If $0 \le p \le 1$, there exist $A, B \subset \mathbf{R}$ of Hausdorff dimension p such that $H_p(A) = \infty$ and $H_p(B) = 0$. (Use Exercise 14.)

16. If $A \subset \mathbf{R}^n$ has Hausdorff dimension p, then $A \times A \subset \mathbf{R}^{2n}$ has Hausdorff dimension $2p$.

3. INTEGRATION ON MANIFOLDS

This section is a brief informal essay on the integration on manifolds for the benefit of those who are familiar with the language of differential geometry and wish to see how the geometric notions of integration fit into the measure-theoretic framework.

Let M be a C^∞ manifold of dimension n. (We work in the C^∞ category for convenience; C^1 would actually suffice, with a few obvious terminological changes.) Given a coordinate system $x = (x_1, \dots, x_n)$ on an open set $U \subset M$, one can consider Lebesgue measure $dx = dx_1 \cdots dx_n$ on U. This has no intrinsic geometric significance, for if $y = (y_1, \dots, y_n)$ is another coordinate system on U, we have $dy = |\det(\partial y/\partial x)|\, dx$, where $(\partial y/\partial x)$ is shorthand for the matrix $(\partial y_i/\partial x_j)$. However, dy and dx are absolutely continuous with respect to each other, and the Radon–Nikodym derivative $|\det(\partial y/\partial x)|$ is a C^∞ function. It therefore makes sense to define a **smooth measure** on M to be a Borel measure μ which, in any local coordinates x, has the form $d\mu = \varphi^x\, dx$ where φ^x is a nonnegative C^∞ function. The representations of μ in different coordinate systems are then related by

$$\varphi^x = |\det(\partial y/\partial x)|\varphi^y. \tag{10.35}$$

[This conveniently sloppy notation is in the same spirit as the formula $du/dt = (du/dx)(dx/dt)$ for the chain rule. More precisely, if $y = f(x)$, then $\varphi^x = |\det(\partial y/\partial x)|(\varphi^y \circ f)$.]

Equation (10.35) may be interpreted in the language of geometry as follows. The functions $|\det(\partial y/\partial x)|$ are the transition functions for a line bundle on M; a section of this bundle, called a **density** on M, is an object which in each local coordinate system x is represented by a smooth function φ^x, such that the functions for different coordinate systems are related by (10.35). In short, smooth measures can be identified with nonnegative densities on M. More generally, any density φ defines, at least locally, a smooth signed or complex measure μ on M, so the integral $\int_K \varphi = \mu(K)$ is well defined for any compact $K \subset M$, as is $\int f\varphi = \int f\, d\mu$ for any $f \in C_c(M)$.

Suppose now that M is equipped with a Riemannian metric. In any coordinate system x, the metric is represented by a positive definite matrix-valued function $g^x = (g^x_{ij})$. The matrix g^y for another coordinate system is related to g^x by

$$g^x_{ij} = \sum_{kl} g^y_{kl} \frac{\partial y_i}{\partial x_k} \frac{\partial y_j}{\partial x_l}, \quad \text{or} \quad g^x = \left(\frac{\partial y}{\partial x}\right)^* g^y \left(\frac{\partial y}{\partial x}\right),$$

so that

$$\det g^x = \left[\det\left(\frac{\partial y}{\partial x}\right)\right]^2 \det g^y.$$

It follows that $\sqrt{\det g}$ is a positive density on M canonically associated to the metric; it is called the **Riemannian volume density** on M. In particular, if M is a submanifold of $\mathbf{R}^m$ $(m \ge n)$, it inherits a Riemannian structure from the ambient Euclidean structure. If M is parametrized by $f\colon V \to \mathbf{R}^m$ as in Section 2, the metric g is given in the coordinates induced by f by

$$g^x_{ij} = \sum_k \frac{\partial f_k}{\partial x_i} \frac{\partial f_k}{\partial x_j}, \quad \text{or} \quad g^x = \left(\frac{\partial f}{\partial x}\right)^* \left(\frac{\partial f}{\partial x}\right).$$

Theorem (10.28) therefore asserts that integration of the Riemannian volume density gives the n-dimensional Hausdorff measure of M (modulo the normalization constant γ_n).

The other popular things to integrate on manifolds are differential forms. For our purposes it will suffice to describe a differential n-form on an n-dimensional manifold M as a section of the line bundle whose transition functions are $\det(\partial y/\partial x)$. That is, a differential n-form ω is given in local coordinates x by a function ω^x such that $\omega^x = \det(\partial y/\partial x)\omega^y$. (The usual notation is $\omega = \omega^x\, dx_1 \wedge \cdots \wedge dx_n$.) Differential n-forms thus look just like densities if one restricts oneself to coordinate systems whose transition matrices $(\partial y/\partial x)$ have positive determinant. If it is possible to do this consistently on all of M, M is called **orientable**. In this case, assuming M is connected, the coordinate systems on subsets of M fall into two classes such that within each class one always has $\det(\partial y/\partial x) > 0$; a choice of one of these classes is an **orientation** of M. If M is equipped with an orientation, therefore, differential n-forms may be identified with densities and as such may be integrated over compact subsets of M.

The notion of density can be generalized. If $0 < \theta \le 1$, a **θ-density** on M is a section of the line bundle whose transition functions are $|\det(\partial y/\partial x)|^\theta$. If φ is a θ-density and $p = \theta^{-1}$, $|\varphi|^p$ is well defined as a nonnegative 1-density and so can be integrated over M. The set of θ-densities φ such that $\|\varphi\|_p = (\int |\varphi|^p)^{1/p} < \infty$ is a normed linear space whose completion $L^p(M)$ is called the **intrinsic L^p space** of M. The duality results of Chapter 6 work in this setting: if φ_j is a θ_j-density for $j = 1, 2$, where $\theta_1 + \theta_2 = 1$, then $\varphi_1\varphi_2$ is a 1-density and $|\int \varphi_1\varphi_2| \le \|\varphi_1\|_{p_1}\|\varphi_2\|_{p_2}$ where $p_j = \theta_j^{-1}$, and $L^{p_1}(M) \cong (L^{p_2}(M))^*$.

4. NOTES AND REFERENCES

Section 1. The existence and uniqueness of Haar measure were first proved by Haar and von Neumann, respectively, for groups whose topology is second countable; the general case is due to Weil. Our proof of existence and uniqueness follows Weil [109] and Loomis [65]. There is another proof, due to H. Cartan, which yields existence and uniqueness simultaneously and avoids the use of the axiom of choice (which we invoked via Tychonoff's theorem). This proof, as well as further references and historical remarks, can be found in Hewitt and Ross [53], Section 15.

Haar measure is the foundation for harmonic analysis on locally compact groups. The articles by Graham, Weiss, and Sally in Ash [4] provide a good introduction to this field, along with references for further reading.

Section 2. Proposition (10.19) is due to Carathéodory [16]. Hausdorff measure was invented by Hausdorff [48], whose proof of Theorem (10.31) we have reproduced. The computation of the normalization constant γ_n can be found in Billingsley [12]. There are other ways of defining lower-dimensional measures on $\mathbf{R}^n$, all of which agree on smooth submanifolds but may differ on more irregular sets; see Federer [29].

The concept of Hausdorff measure can be generalized. If λ is any strictly increasing continuous function on $[0, \infty)$ such that $\lambda(0) = 0$, and if A is a subset of a metric space X, one can define

$$H_{\lambda,\delta}(A) = \inf\left\{\sum_1^\infty \lambda(\operatorname{diam} B_j) \colon A \subset \bigcup_1^\infty B_j, \operatorname{diam} B_j \le \delta\right\}$$

and $H_\lambda(A) = \lim_{\delta \to 0} H_{\lambda,\delta}(A)$. Thus $H_p = H_\lambda$ where $\lambda(t) = t^p$. The basic properties of H_p generalize to H_λ, and if λ is concave, the proof of Theorem (10.31), with $\xi_n = \lambda^{-1}(2^{-n})$, yields a subset B of $[0, \xi_0]$ such that $H_\lambda(B) = 1$. See Rogers [85] for more about abstract Hausdorff measures.

Continuous curves of Hausdorff dimension greater than one can be constructed from nondifferentiable functions. For example, if $f\colon [a, b] \to \mathbf{R}$ is Hölder continuous of exponent α, $0 < \alpha < 1$, the graph of f in $\mathbf{R}^2$ may have Hausdorff dimension as large as $2 - \alpha$; and the range of a sample path of the n-dimensional Wiener process, $n \ge 2$, almost surely has Hausdorff dimension 2. See Besicovitch and Ursell [11] and Taylor [106]. The geometry of sets of nonintegral Hausdorff dimension is a subject of some fascination; see Mandelbrot [67].

Section 3. A more complete treatment of integration on manifolds can be found, for example, in Loomis and Sternberg [66].

BIBLIOGRAPHY

1. W. J. Adams, *The Life and Times of the Central Limit Theorem*, Kaedmon, New York, 1974.
2. L. Alaoglu, Weak convergence of linear functionals, *Bull. Amer. Math. Soc.* **44** (1938), 196.
3. L. Alaoglu, Weak topologies of normed linear spaces, *Annals of Math. (2)* **41** (1940), 252–267.
4. J. M. Ash (ed.), *Studies in Harmonic Analysis*, Mathematical Association of America, Washington, D.C., 1976.
5. S. Banach, Sur le problème de la mesure, *Fund. Math.* **4** (1923), 7–33.
6. S. Banach, *Théorie des Opérations Linéaires*, Monografje Matematyczne, Warsaw, 1932.
7. S. Banach and H. Steinhaus, Sur le principe de la condensation de singularités, *Fund. Math.* **9** (1927), 50–61.
8. S. Banach and A. Tarski, Sur la décomposition des ensembles de points en parties respectivement congruentes, *Fund. Math.* **6** (1924), 244–277.
9. R. G. Bartle, An extension of Egorov's theorem, *Amer. Math. Monthly* **87** (1980), 628–633.
10. J. Bergh and J. Löfström, *Interpolation Spaces*, Springer-Verlag, Berlin, 1976.
11. A. S. Besicovitch and H. D. Ursell, Sets of fractional dimension (V): on dimensional numbers of some continuous curves, *J. London Math. Soc.* **12** (1937), 18–25.
12. P. Billingsley, *Probability and Measure*, Wiley, New York, 1979.
13. N. Bourbaki, Sur les espaces de Banach, *C. R. Acad. Sci. Paris* **206** (1938), 1701–1704.
14. N. Bourbaki, *General Topology* (2 vols.), Hermann, Paris, and Addison-Wesley, Reading, Mass., 1966.
15. A. Brown, An elementary example of a continuous singular function, *Amer. Math. Monthly* **76** (1969), 295–297.
16. C. Carathéodory, *Vorlesungen über Reelle Funktionen*, Teubner, Leipzig, 1918; 2nd ed., 1927, reprinted by Chelsea, New York, 1948.
17. K. L. Chung, *A Course in Probability Theory*, 2nd ed., Academic Press, New York, 1974.
18. P. J. Cohen, *Set Theory and the Continuum Hypothesis*, Benjamin, New York, 1966.
19. D. L. Cohn, *Measure Theory*, Birkhäuser, Boston, 1980.
20. R. Coifman, M. Cwikel, R. Rochberg, Y. Sagher, and G. Weiss, Complex interpolation for families of Banach spaces, in *Harmonic Analysis in Euclidean Spaces*, Proc. Symp. Pure Math., Vol. 35, part 2, American Mathematical Society, Providence, R.I., 1979, pp. 269–282.
21. P. J. Daniell, A general form of integral, *Annals of Math. (2)* **19** (1918), 279–294.
22. K. deLeeuw, The Fubini theorem and convolution formula for regular measures, *Math. Scand.* **11** (1962), 117–122.

23. J. Dieudonné, *History of Functional Analysis*, North-Holland, Amsterdam, 1981.
24. R. Doss, The Hahn decomposition theorem, *Proc. Amer. Math. Soc.* **80** (1980), 377.
25. J. Dugundji, *Topology*, Allyn and Bacon, Boston, 1966.
26. N. Dunford and J. T. Schwartz, *Linear Operators* (3 vols.), Wiley-Interscience, New York, 1958, 1963, and 1971.
27. H. Dym and H. P. McKean, *Fourier Series and Integrals*, Academic Press, New York, 1972.
28. R. Engelking, *General Topology*, PWN—Polish Scientific Publishers, Warsaw, 1977.
29. H. Federer, *Geometric Measure Theory*, Springer-Verlag, Berlin, 1969.
30. M. B. Feldman, A proof of Lusin's theorem, *Amer. Math. Monthly* **88** (1981), 191–192.
31. E. Fischer, Sur le convergence en moyenne, *C. R. Acad. Sci. Paris* **144** (1907), 1022–1024.
32. G. B. Folland, *Introduction to Partial Differential Equations*, Princeton University Press, Princeton, N.J., 1976.
33. G. B. Folland, *Lectures on Partial Differential Equations*, Tata Institute of Fundamental Research Lectures No. 70, Springer-Verlag, Berlin, 1983.
34. G. B. Folland and E. M. Stein, Estimates for the $\overline{\partial}_b$ complex and analysis on the Heisenberg group, *Comm. Pure Appl. Math.* **27** (1974), 429–522.
35. M. Fréchet, Sur quelques points du calcul fonctionnel, *Rend. Circ. Mat. Palermo* **22** (1906), 1–74.
36. M. Fréchet, Sur l'intégrale d'une fonctionnelle étendue à un ensemble abstrait, *Bull. Soc. Math. France* **43** (1915), 248–265.
37. M. Fréchet, Des familles et fonctions additives d'ensembles abstraits, *Fund. Math.* **5** (1924), 206–251.
38. K. Gödel, What is Cantor's continuum problem?, *Amer. Math. Monthly* **54** (1947), 515–525. A revised and expanded version appears in P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics*, Prentice-Hall, Englewood Cliffs, N.J., 1964.
39. H. Hahn, Über die Multiplikation total-additiver Mengenfunktionen, *Annali Scuola Norm. Sup. Pisa* **2** (1933), 429–452.
40. H. Hahn and A. Rosenthal, *Set Functions*, Univ. of New Mexico Press, Albuquerque, N.M., 1948.
41. P. R. Halmos, *Measure Theory*, Van Nostrand, Princeton, N.J., 1950.
42. P. R. Halmos, *Naive Set Theory*, Van Nostrand, Princeton, N.J., 1960.
43. G. H. Hardy and J. E. Littlewood, Some properties of fractional integrals I, *Math. Zeit.* **27** (1927), 565–606.
44. G. H. Hardy and J. E. Littlewood, A maximal theorem with function-theoretic applications, *Acta Math.* **54** (1930), 81–116.
45. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge University Press, Cambridge, U.K., 1934.
46. D. G. Hartig, The Riesz representation theorem revisited, *Amer. Math. Monthly* **90** (1983), 277–280.
47. F. Hausdorff, *Grundzüge der Mengenlehre*, Verlag von Veit, Leipzig, 1914; reprinted by Chelsea, New York, 1949.
48. F. Hausdorff, Dimension und äusseres Mass, *Math. Annalen* **79** (1919), 157–179.
49. T. Hawkins, *Lebesgue's Theory of Integration*, Univ. of Wisconsin Press, Madison, Wisc., 1970.
50. J. Hennefeld, A nontopological proof of the uniform boundedness theorem, *Amer. Math. Monthly* **87** (1980), 217.
51. E. Hewitt, On two problems of Urysohn, *Annals of Math. (2)* **47** (1946), 503–509.

52. E. Hewitt and R. E. Hewitt, The Gibbs-Wilbraham phenomenon: an episode in Fourier analysis, *Arch. Hist. Exact Sci.* **21** (1979), 129–160.
53. E. Hewitt and K. A. Ross, *Abstract Harmonic Analysis*, Vol. I, Springer-Verlag, Berlin, 1963.
54. E. Hewitt and K. Stromberg, *Real and Abstract Analysis*, Springer-Verlag, Berlin, 1965.
55. S. Kakutani, Concrete representation of abstract (M)-spaces, *Annals of Math. (2)* **42** (1941), 994–1024.
56. S. Kakutani and J. C. Oxtoby, Construction of a non-separable invariant extension of the Lebesgue measure space, *Annals of Math. (2)* **52** (1950), 580–590.
57. J. L. Kelley, The Tychonoff product theorem implies the axiom of choice, *Fund. Math.* **37** (1950), 75–76.
58. J. L. Kelley, *General Topology*, Van Nostrand, Princeton, N.J., 1955.
59. A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer-Verlag, Berlin, 1933; translated as *Foundations of the Theory of Probability*, Chelsea, New York, 1950.
60. A. V. Lair, A Rellich compactness theorem for sets of finite volume, *Amer. Math. Monthly* **83** (1976), 350–351.
61. J. Lamperti, *Probability*, Benjamin, New York, 1966.
62. H. Lebesgue, Intégrale, longueur, aire, *Annali Mat. Pura Appl. (3)* **7** (1902), 231–359.
63. H. Lebesgue, Sur l'intégration des fonctions discontinues, *Ann. Sci. Ecole Norm. Sup.* **27** (1910), 361–450.
64. H. Lebesgue, *Leçons sur l'Intégration*, 2nd ed., Gauthier-Villars, Paris, 1928.
65. L. H. Loomis, *An Introduction to Abstract Harmonic Analysis*, Van Nostrand, Princeton, N.J., 1953.
66. L. H. Loomis and S. Sternberg, *Advanced Calculus*, Addison-Wesley, Reading, Mass., 1968.
67. B. B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, 1982.
68. J. Marcinkiewicz, Sur l'interpolation d'opérateurs, *C. R. Acad. Sci. Paris* **208** (1939), 1272–1273.
69. J. Nagata, *Modern General Topology*, North-Holland, Amsterdam, 1968.
70. E. Nelson, Regular probability measures on function spaces, *Annals of Math. (2)* **69** (1959), 630–643.
71. E. Nelson, Feynman integrals and the Schrödinger equation, *J. Math. Phys.* **5** (1964), 332–343.
72. E. Nelson, *Dynamical Theories of Brownian Motion*, Princeton Univ. Press, Princeton, N.J., 1967.
73. J. von Neumann, *Mathematische Grundlagen der Quantenmechanik*, Springer-Verlag, Berlin, 1932.
74. D. J. Newman, Fourier uniqueness via complex variables, *Amer. Math. Monthly* **81** (1974), 379–380.
75. O. Nikodym, Sur une généralisation des intégrales de M. J. Radon, *Fund. Math.* **15** (1930), 131–179.
76. W. F. Pfeffer, *Integrals and Measures*, Marcel Dekker, New York, 1977.
77. J. Radon, Theorie und Anwendungen der absolut additiv Mengenfunktionen, *S.-B. Math.-Natur. Kl. Kais. Akad. Wiss. Wien* **122**.IIa (1913), 1295–1438.
78. M. Reed and B. Simon, *Methods of Mathematical Physics I: Functional Analysis*, Academic Press, New York, 1972.
79. B. Riemann, Über die Hypothesen, welche der Geometrie zu Grunde liegen, in *Gesammelte Mathematische Werke*, Teubner, Leipzig, 1876, pp. 254–269.
80. F. Riesz, sur les systèmes orthogonaux de fonctions, *C. R. Acad. Sci. Paris* **144** (1907), 615–619.

81. F. Riesz, Sur une espèce de géométrie analytique des systèmes de fonctions sommables, *C. R. Acad. Sci. Paris* **144** (1907), 1409–1411.
82. F. Riesz, Sur les opérations fonctionnelles linéaires, *C. R. Acad. Sci. Paris* **149** (1909), 974–977.
83. F. Riesz, Untersuchungen über Systeme integrierbarer Funktionen, *Math. Annalen* **69** (1910), 449–497.
84. M. Riesz, Sur les maxima des formes bilinéaires et sur les fonctionnelles linéaires, *Acta Math.* **49** (1926), 465–497.
85. C. A. Rogers, *Hausdorff Measures*, Cambridge Univ. Press, Cambridge, U.K., 1970.
86. J. L. Romero, When is $L^p(\mu)$ contained in $L^q(\mu)$?, *Amer. Math. Monthly* **90** (1983), 203–206.
87. H. L. Royden, *Real Analysis*, 2nd ed., Macmillan, New York, 1968.
88. L. A. Rubel, A complex-variables proof of Hölder's inequality, *Proc. Amer. Math. Soc.* **15** (1964), 999.
89. W. Rudin, *Functional Analysis*, McGraw-Hill, New York, 1973.
90. W. Rudin, *Real and Complex Analysis*, 2nd ed., McGraw-Hill, New York, 1974.
91. W. Rudin, Lebesgue's first theorem, in L. Nachbin (ed.), *Mathematical Analysis and Applications*, (Advances in Math. Supplementary Studies, Vol. 7B), Academic Press, New York, 1981, pp. 741–747.
92. W. Rudin, Well-distributed measurable sets, *Amer. Math. Monthly* **90** (1983), 41–42.
93. S. Saks, *Theory of the Integral*, 2nd ed., Monografje Matematyczne, Warsaw, 1937; reprinted by Hafner, New York, 1938.
94. J. Schwartz, A note on the space L_p^*, *Proc. Amer. Math. Soc.* **2** (1951), 270–275.
95. J. Schwartz, The formula for change in variables in a multiple integral, *Amer. Math. Monthly* **61** (1954), 81–85.
96. L. Schwartz, *Théorie des Distributions*, 2nd ed., Hermann, Paris, 1966.
97. J. Serrin and D. E. Varberg, A general chain rule for derivatives and the change of variables formula for the Lebesgue integral, *Amer. Math. Monthly* **76** (1969), 514–520.
98. W. Sierpiński, Sur un problème concernant les ensembles measurables superficiellement, *Fund. Math.* **1** (1920), 112–115.
99. R. M. Solovay, A model of set theory in which every set of reals is Lebesgue measurable, *Annals of Math. (2)* **92** (1970), 1–56.
100. E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton Univ. Press, Princeton, N.J., 1970.
101. E. M. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton Univ. Press, Princeton, N.J., 1971.
102. H. Steinhaus, Additive und stetige Funktionaloperationen, *Math. Zeit.* **5** (1919), 186–221.
103. M. H. Stone, The generalized Weierstrass approximation theorem, *Math. Mag.* **21** (1948), 167–184 and 237–254; reprinted in R. C. Buck (ed.), *Studies in Modern Analysis*, Mathematical Association of America, Washington, D.C., 1962.
104. K. Stromberg, The Banach-Tarski paradox, *Amer. Math. Monthly* **86** (1979), 151–161.
105. M. E. Taylor, *Pseudodifferential Operators*, Princeton University Press, Princeton, N.J., 1981.
106. S. J. Taylor, The Hausdorff α-dimensional measure of Brownian paths in n-space, *Proc. Cambridge Phil. Soc.* **49** (1953), 31–39.
107. G. O. Thorin, An extension of a convexity theorem due to M. Riesz, *Kungl. Fysiografiska Saellskapet i Lund Forhaendlinger* **8** (1939), no. 14.

108. F. Treves, *Topological Vector Spaces, Distributions, and Kernels*, Academic Press, New York, 1967.
109. A. Weil, *L'Intégration dans les Groupes Topologiques et ses Applications*, Hermann, Paris, 1940.
110. N. Wiener, The ergodic theorem, *Duke Math. J.* **5** (1939), 1–18.
111. C. S. Wong, A note on the central limit theorem, *Amer. Math. Monthly* **84** (1977), 472.
112. A. Zygmund, On a theorem of Marcinkiewicz concerning interpolation of operators, *J. Math. Pures Appl. (9)* **35** (1956), 223–248.
113. A. Zygmund, *Trigonometric Series* (2 vols., reprinted in 1 vol.), Cambridge Univ. Press, Cambridge, U.K., 1968.

INDEX

Abel mean, 251
Absolute convergence, 144
Absolutely continuous function, 100
Absolutely continuous measure, 83
Abstract Wiener process, 305
Accumulation point, 108
Additivity, countable, 24
  finite, 24
Adjoint, 152, 169
A.e., 26
Alaoglu's theorem, 162
Alexander's lemma, 130
Alexandrov compactification, 126
Algebra, Banach, 146
  of functions, 133
  of sets, 20
Almost disjoint cubes, 71
Almost everywhere, 26
Almost surely, 287
Almost uniform convergence, 60
Approximate identity, 236
Arcwise connected set, 118
A.s., 287
Ascoli-Arzelà theorem, 131
Axiom, of choice, 6
  of countability, 109
  of separation, 110

Baire category theorem, 153
Baire class, 79
Baire set, 208
Ball, 13
Banach algebra, 146
Banach space, 144
Base, 108
  neighborhood, 108
Basis, orthonormal, 168
Bernstein polynomial, 291
Bernstein's theorem, 253
Bessel's inequality, 167
Bijective mapping, 4
Binomial distribution, 293
Bochner integral, 171
Bolzano-Weierstrass property, 15
Borel-Cantelli lemma, 294
Borel measurable function, 43
Borel set, 21
Borel σ-algebra, 21
Borel's normal number theorem, 304
Boundary, 108
Bounded linear map, 145
Bounded set, 15
Bounded variation, function of, 97

Cantor function, 38
Cantor-Lebesgue function, 38
Cantor set, 37
  generalized, 41
Carathéodory's theorem, 28
Cardinality of the continuum, 8
Cartesian product, 3, 4
Category, first, 154
  second, 154
Cauchy net, 158
Cauchy-Riemann equation, 279
Cauchy sequence, 14
  in measure, 59
Central limit theorem, 300
Cesàro mean, 252
Characteristic function, 45, 287
Chebyshev's inequality, 185
Chi-square distribution, 293
Choice, axiom of, 6
Class, Baire, 79
  monotone, 64
Closed graph theorem, 155

Closed linear map, 155
Closed set, 13, 108
Closure, 14, 108
Closure operator, 112
Cluster point, of a net, 120
  of a sequence, 112
Coarser topology, 108
Co-countable set, 20
Cofinite topology, 107
Commutator subgroup, 320
Compactification, 126, 138
  Alexandrov, 126
  one-point, 126
  Stone-Čech, 138
Compactly supported function, 125
Compact space, 16, 122
Complement, 3
Completely regular algebra, 140
Completely regular space, 117
Complete measure, 26
Complete metric space, 14
Complete orthonormal set, 168
Complete topological vector space, 158
Completion, of a measure, 26
  of a normed vector space, 151
  of a σ-algebra, 26
Complex Hahn-Banach theorem, 150
Complex measure, 88
Complex Radon measure, 216
Complex Stone-Weierstrass theorem, 135
Composition, 3
Condensation of singularities, 158
Conditional expectation, 88
Conjugate exponent, 175
Connected component, 112
Connected set, 112
Constant-coefficient operator, 275
Continuity, uniform, 14, 229, 313
Continuous mapping, 14, 113
Continuous measure, 102
Continuum, cardinality of, 8
Continuum hypothesis, 17
Convergence, absolute, 144
  almost uniform, 60
  of a filter, 141
  in measure, 59
  of a net, 120
  in probability, 287
  of a sequence, 14, 110
  of a series, 144
  weak, 161
Converse of Hölder's inequality, 181
Convex function, 104
Convolution, of distributions, 260, 263, 267
  of functions, 230
  of measures, 281
Coordinate, 4
Coordinate map, 4
Countability, axioms of, 109
Countable additivity, 24
Countable ordinals, 10
Countable set, 7
Countably compact space, 124
Counting measures, 25
Cover, 15
  locally finite, 129
Cube, 71, 137
Cumulative distribution function, 34

Daniell integral, 78
Decomposable measure, 87
Decomposition, Hahn, 82
  Lebesgue, 85
Decreasing function, 12
Decreasing rearrangement, 192
DeMorgan's laws, 3
Denominator, 329
Dense set, 14, 108
Density, 332
  probability, 293
  Riemannian volume, 333
  of a set, 95
Derivative, of a distribution, 260, 263
  L^p, 237
  Radon-Nikodym, 85
Diameter, 15
Diffeomorphism, 71
Difference of sets, 3
Differential operator, 274
  elliptic, 278, 285
Dimension, Hausdorff, 325
Dirac measure, 25
Directed set, 119
Dirichlet kernel, 253
Dirichlet problem, 276
Disconnected set, 112
Discrete measure, 102
Discrete topology, 107
Disjoint sets, 2
Distribution, 262, 288
  binomial, 293
  chi-square, 293
  gamma, 293
  Gaussian, 299
  harmonic, 265
  joint, 288
  normal, 299
  Poisson, 293

  tempered, 258
Distribution function, 34, 189, 288
Domain of a mapping, 4
Dominated convergence theorem, 53
Dual space, 148

Egoroff's theorem, 60
Elementary family, 22
Elliptic differential operator, 278, 285
Elliptic regularity theorem, 278
Equicontinuous family, 131
Equivalence class, 3
Equivalence relation, 3
Equivalent metrics, 17
Equivalent norms, 143
Ergodic theorem, mean, 171
Essential range, 179
Essential supremum, 176
Event, 286
Events, independent, 288
Eventually, 120
Expectation, 287
  conditional, 88
Expected value, 287
Exponent, conjugate, 175
Extended integrable function, 81
Extended real number system, 11

Family, elementary, 22
Fatou's lemma, 50
Fejér kernel, 246, 257
Field of sets, 20
Filter, 141
Finer topology, 108
Finite intersection property, 122
Finitely additive measure, 24
Finite measure, 24
Finite part of divergent integral, 267
Finite sequence, 4
Finite signed measure, 83
First category, 154
First countable space, 109
Formula, Poisson summation, 249
  Stirling's, 301
Fourier inversion theorem, 243
Fourier-Stieltjes transform, 283
Fourier transform, 240, 260
Fractional integral, 78
Fréchet space, 158
Frequently, 120
F_σ and $F_{\sigma\delta}$ sets, 21
Fubini's theorem, 65
Fubini-Tonelli theorem, 65, 223
Function, 3
  absolutely continuous, 100
  of bounded variation, 97
  Cantor, 38
  characteristic, 45, 287
  compactly supported, 125
  convex, 104
  decreasing, 12
  distribution, 34, 189, 288
  extended integrable, 81
  Hermite, 248
  Hölder continuous, 132, 147, 252
  increasing, 12
  Lipschitz continuous, 103
  lower semi-continuous, 211
  l.s.c., 211
  maximal, 91, 238
  measurable, 43
  modular, 320
  monotone, 12
  piecewise regular, 256
  periodic, 229
  real analytic, 258
  right continuous, 33
  simple, 45
  slowly increasing, 259
  tempered, 258
  upper semi-continuous, 211
  u.s.c., 211
Functional, linear, 148
  Minkowski, 149
Fundamental theorem of calculus, 102

Gamma distribution, 293
Gaussian distribution, 299
Gaussian kernel, 246
G_δ and $G_{\delta\sigma}$ sets, 21
Generalized Cantor set, 41
Gibbs phenomenon, 256
Gram-Schmidt process, 167
Graph of a linear map, 155
Group, locally compact, 314
  topological, 312
  unimodular, 320

Haar measure, 314
Hahn-Banach theorem, 149, 150
Hahn decomposition, 82
Hahn decomposition theorem, 81
Hardy-Littlewood maximal function, 91
Hardy's inequalities, 188
Harmonic distribution, 265
Hausdorff dimension, 325
Hausdorff maximal principle, 5
Hausdorff measure, 323

Hausdorff space, 110
Hausdorff-Young inequality, 240, 245
Heat equation, 277
Heine-Borel property, 15
Heisenberg's inequality, 189
Hermite function, 248
Hermite operator, 248
Hermite polynomial, 248
Hilbert's inequality, 188
Hilbert space, 164
Hilbert transform, 266
H-interval, 32
Hölder continuous function, 132, 147, 252
Hölder's inequality, 174
  converse of, 181
  generalized, 189
Homeomorphism, 113
Hull of an ideal, 136

Ideal, 136
Identically distributed random variables, 288
Identity, approximate, 236
Image, 3
  inverse, 3
Image measure, 287
Inclusion, reverse, 119
Increasing function, 12
Independent events, 288
Independent random variables, 288
Indicator function, 45, 287
Inequality, Bessel's, 167
  Chebyshev's, 185
  Hardy's, 188
  Hausdorff-Young, 240, 245
  Heisenberg's, 189
  Hilbert's, 188
  Hölder's, 174, 189
  Jensen's, 104
  Kolmogorov's, 295
  Minkowski's, 175, 186
  Schwarz, 164
  triangle, 143
  Wirtinger's, 247
  Young's, 232
Infimum, 9, 11
Initial segment, 9
Injective mapping, 4
Inner measure, 31
Inner product, 163
Inner regular measure, 205
Integrable function, 51
  extended, 81
  locally, 90
  Riemann, 55
Integral, Bochner, 171
  of a complex function, 51
  Daniell, 78
  fractional, 78
  Lebesgue, 55, 68
  of a nonnegative function, 48
  Riemann, 55
  of a simple function, 47
Interior, 13, 108
Intrinsic L^p space, 333
Inverse, 4
Inverse image, 3
Invertible linear map, 146
Isometry, 146
Isomorphism, 146
  unitary, 169

Jensen's inequality, 104
Joint distribution, 288
Jordan decomposition, of a function, 98
  of a measure, 82

Kernel, of a closed set, 136
  Dirichlet, 253
  Fejér, 246, 257
  Gaussian, 246
  Poisson, 246, 256
  Weierstrass, 246
Khinchine's strong law of large numbers, 296
Kolmogorov's inequality, 295
Kolmogorov's strong law of large numbers, 295
Krein's extension theorem, 153
Kuratowski closure operator, 112

Laplacian, 275
Lattice of functions, 133
Law, deMorgan's, 3
  of large numbers, 294 ff.
LCH space, 124
Lebesgue decomposition, 85
Lebesgue differentiation theorem, 93
Lebesgue integral, 55, 68
Lebesgue measurable function, 43
Lebesgue measurable set, 37, 68
Lebesgue measure, 37, 68
Lebesgue-Radon-Nikodym theorem, 84, 88
Lebesgue set, 92
Lebesgue-Stieltjes integral, 100
Lebesgue-Stieltjes measure, 35
Lemma, Alexander's, 130
  Borel-Cantelli, 294
  Fatou's, 50

  Riemann-Lebesgue, 241
  three lines, 193
  Urysohn's, 116, 125, 237
  Weyl's, 279
  Zorn's, 5
Left Haar measure, 314
Left uniform continuity, 313
Limit inferior, 2, 11
Limit of a net, 120
Limit superior, 2, 11
Linear functional, 148
  positive, 204
Linearly ordered set, 5
Linear map, bounded, 145
  closed, 155
  invertible, 146
Lipschitz continuous function, 103
Localization principle, 254
Locally compact group, 314
Locally compact space, 124
Locally convex space, 158
Locally finite cover, 129
Locally integrable function, 90
Locally measurable set, 27
Locally null set, 184
Lower bound, 5
Lower semi-continuous function, 211
L^p derivative, 237
L^p space, 52, 173, 176
  intrinsic, 333
  weak, 191
L.s.c. function, 211
Lusin's theorem, 61, 211

Mapping, 3
  bijective, 4
  continuous, 14, 113
  injective, 4
  measurable, 42
  open, 154
  proper, 128
  sublinear, 195
  surjective, 4
  uniformly continuous, 14
Marcinkiewicz interpolation theorem, 195
Maximal element, 5
Maximal function, 91, 238
Maximal theorem, 91
Meager set, 154
Mean, 287, 293
  Abel, 251
  Cesàro, 252
  sample, 298
Mean ergodic theorem, 171
Measurable function, 43
Measurable mapping, 42
Measurable set, 24
  Lebesgue, 37, 68
  with respect to an outer measure, 28
Measurable space, 24
Measure, 24
  absolutely continuous, 83
  Cauchy sequence in, 59
  complex, 88
  complex Radon, 216
  continuous, 102
  convergence in, 59
  counting, 25
  Dirac, 25
  discrete, 102
  finite, 24
  finitely additive, 24
  Hausdorff, 323
  image, 287
  inner, 31
  inner regular, 205
  Lebesgue, 37, 68
  Lebesgue-Stieltjes, 35
  metric outer, 322
  outer, 28
  outer regular, 205
  positive, 80
  probability, 286
  Radon, 205
  saturated, 27
  σ-finite, 24
  signed, 80
  signed Radon, 216
  singular, 82
  smooth, 332
  tempered, 259
Measure space, 24
Metric, 13
  equivalent, 17
  product, 13
Metric outer measure, 322
Metric space, 13
  complete, 14
  separable, 14
Minimal element, 5
Minkowski functional, 149
Minkowski's inequality, 175
  for integrals, 186
Modular function, 320
Moment convergence theorem, 301
Monotone class, 64
Monotone convergence theorem, 49

Monotone function, 12
Monotonicity of measures, 25
Multi-index, 226
Mutually singular measures, 82

Negative part of a function, 45
Negative set, 81
Negative variation, of a function, 98
  of a measure, 82
Neighborhood, 108
Neighborhood base, 108
Net, 119
  Cauchy, 158
Norm, 143
  equivalent, 143
  L^P, 173
  operator, 145
  product, 145
  quotient, 147
  uniform, 115
Normal distribution, 299
Normal number, 304
Normal space, 110
Normed linear space, 143
Normed vector space, 143
Norm topology, 143
Nowhere dense set, 14, 108
Null set, 26, 81
Number, normal, 304

One-point compactification, 126
Open mapping, 154
Open mapping theorem, 154
Open set, 12, 13, 108
Operator, constant-coefficient, 275
  differential, 274
  elliptic, 278, 285
  Hermite, 248
  Laplace, 275
Operator norm, 145
Operator topology, strong, 161
  weak, 161
Ordering, 5
  linear, 5
  partial, 5
  total, 5
  well, 5
Order isomorphism, 5
Order topology, 112
Ordinal, countable, 10
  first uncountable, 10
Orientable manifold, 333
Orientation, 333
Orthogonal projection, 170
Orthogonal vectors, 165
Orthonormal basis, 168
Orthonormal set, 167
Outer measure, 28
  metric, 322
Outer regular measure, 205

Paracompact space, 129
Parametrization, 325
Parallelogram law, 165
Parseval's equation, 168
Partial ordering, 5
Partition, of an interval, 55
  of unity, 127
Periodic function, 229
Piecewise regular function, 256
Plancherel theorem, 244
Point mass, 25
Pointwise bounded family, 131
Poisson distribution, 293
Poisson kernel, 246, 256
Poisson summation formula, 249
Polar decomposition, 45
Polynomial, Bernstein, 291
  Hermite, 248
Positive linear functional, 204
Positive measure, 80
Positive part of a function, 45
Positive set, 81
Positive variation, of a function, 98
  of a measure, 82
Precompact set, 122
Predecessor, 9
Pre-Hilbert space, 164
Premeasure, 30
Principal symbol, 275
Principle, of condensation of singularities, 158
  Hausdorff maximal, 5
  localization, 254
  of transfinite induction, 9
  uniform boundedness, 156
  well ordering, 5
Probability, convergence in, 287
Probability measure, 286
Probability space, 286
Product, Cartesian, 3, 4
  inner, 163
  of measures, 62
  Radon, 221
  scalar, 163
Product metric, 13
Product norm, 145
Product σ-algebra, 22

Product topology, 114
Projection, 4
  orthogonal, 170
Proper map, 128
Pythagorean theorem, 165

Quotient norm, 147
Quotient topology, 118

Radon measure, 205
  complex, 216
  signed, 216
Radon-Nikodym derivative, 85
Radon-Nikodym theorem, 85
Radon product, 221
Random variables, 286
  identically distributed, 288
  independent, 288
Range, 4
  essential, 179
Real analytic function, 258
Rearrangement, decreasing, 192
Rectangle, 62
Refinement of a cover, 129
Reflexive space, 151
Regular measure, 93, 205
Regular space, 110
Relation, 3
  equivalence, 3
Relatively closed set, 108
Relatively open set, 108
Relative topology, 107
Rellich's theorem, 272
Renormalization, 267
Residual set, 154
Reverse inclusion, 119
Riemannian volume density, 333
Riemann integrable function, 55
Riemann integral, 55
Riemann-Lebesgue lemma, 241
Riesz representation theorem, 205, 216
Riesz-Thorin interpolation theorem, 193
Right continuous function, 33
Right Haar measure, 314
Right uniform continuity, 313
Ring of sets, 23

Sample mean, 298
Sample space, 286
Sample variance, 298
Saturated measure, 27
Saturation, 27
Scalar product, 163
Schröder-Bernstein theorem, 7
Schwartz space, 227
Schwarz inequality, 164
Second axiom of countability, 109
Second category, 154
Second countable space, 109
Section of a set or function, 63
Segment, initial, 9
Semi-continuous function, 211
Semifinite measure, 24
Semifinite part, 27
Seminorm, 143
Separable space, 14, 109
Separation axioms, 110, 117
Separation of points, 133
Separation of points and closed sets, 137
Sequence, 4
  Cauchy, 14
  finite, 4
Sequentially compact space, 124
Series, absolutely convergent, 144
  convergent, 144
Set, arcwise connected, 118
  Baire, 208
  Borel, 21
  Cantor, 37, 41
  closed, 13, 108
  connected, 112
  directed, 119
  F_σ, 21
  G_δ, 21
  Lebesgue, 92
  Lebesgue measurable, 37, 68
  linearly ordered, 5
  locally measurable, 27
  locally null, 184
  meager, 154
  measurable, 24
  negative, 81
  null, 26, 81
  open, 12, 13, 108
  orthonormal, 167
  partially ordered, 5
  positive, 81
  precompact, 122
  relatively closed, 108
  relatively open, 108
  residual, 154
  σ-compact, 127
  σ-finite, 24
  totally ordered, 5
  well ordered, 5
Shannon's theorem, 298
Shrink nicely, 93
Side of a rectangle, 68

σ-algebra, 20
  Baire, 208
  Borel, 21
  of countable or co-countable sets, 20
  generated by a family of functions, 43
  generated by a family of sets, 21
  product, 22
σ-compact space, 127
σ-field, 20, 286
σ-finite measure, 24
σ-finite set, 24
σ-finite signed measure, 83
σ-ring, 23
Signed measure, 80
Signed Radon measure, 216
Simple function, 45
Singular measure, 82
Slowly increasing function, 259
Smooth measure, 332
Sobolev embedding theorem, 270, 273
Sobolev space, 268
Space, Banach, 144
  compact, 16, 122
  completely regular, 117
  countably compact, 124
  dual, 148
  Fréchet, 158
  Hausdorff, 110
  Hilbert, 164
  LCH, 124
  locally compact, 124
  locally convex, 158
  L^P, 52, 173, 176
  measurable, 24
  metric, 13
  normal, 110
  normed vector, 143
  paracompact, 129
  pre-Hilbert, 164
  probability, 286
  reflexive, 151
  regular, 110
  sample, 286
  Schwartz, 227
  sequentially compact, 124
  σ-compact, 127
  Sobolev, 268
  T_0, (T_1, etc.), 110, 117
  topological, 107
  topological vector, 158
  Tychonoff, 117
  weak L^P, 191
Standard deviation, 287
Standard normal distribution, 299
Standard representation of a simple function, 45
Stirling's formula, 301
Stone-Čech compactification, 138
Stone-Weierstrass theorem, 133, 135
Strictly decreasing function, 12
Strictly increasing function, 12
Stronger topology, 108
Strong law of large numbers, 295, 296
Strong L^P derivative, 237
Strong operator topology, 161
Strong type, 195
Subadditivity, 25
Subbase, 108
Subgroup, commutator, 320
Sublinear mapping, 195
Submanifold, 325
Subnet, 120
Subordination to a cover, 127
Subsequence, 4
Subspace of a vector space, 143
Support, of a distribution, 263
  of a function, 125
  of a measure, 208
Supremum, 9, 11
  essential, 176
Surjective mapping, 4
Symbol, 275
  principal, 275
Symmetric difference, 3
Symmetric neighborhood, 312

Tempered distribution, 258
Tempered function, 258
Tempered measure, 259
Theorem, Alaoglu's, 162
  Ascoli-Arzelà, 131
  Baire category, 153
  Bernstein's, 253
  Borel's normal number, 304
  Carathéodory's, 28
  central limit, 300
  closed graph, 155
  dominated convergence, 53
  Egoroff's, 60
  elliptic regularity, 278
  Fourier inversion, 243
  Fubini's, 65
  Fubini-Tonelli, 65, 223
  Hahn-Banach, 149, 150
  Hahn decomposition, 81
  Krein's extension, 153
  Lebesgue differentiation, 93
  Lebesgue-Radon-Nikodym, 84, 88

  Lusin's, 61, 211
  Marcinkiewicz interpolation, 195
  maximal, 91
  mean ergodic, 171
  moment convergence, 301
  monotone convergence, 49
  open mapping, 154
  Plancherel, 244
  Pythagorean, 165
  Radon-Nikodym, 85
  Riesz representation, 205, 216
  Riesz-Thorin interpolation, 193
  Rellich's, 272
  Schröder-Bernstein, 7
  Shannon's, 298
  Sobolev embedding, 270, 273
  Stone-Weierstrass, 133, 135
  Tietze extension, 116, 125
  Tonelli's, 65
  Tychonoff's, 130
  Urysohn metrization, 139
  Vitali convergence, 180
  Vitali covering, 105
  Weierstrass approximation, 135, 291
θ-density, 333
Three lines lemma, 193
Tietze extension theorem, 116, 125
T_0 (T_1, etc.) space, 110, 117
Tonelli's theorem, 65
Topological group, 312
Topological space, 107
Topological vector space, 158
Topology, 107
  coarser, 108
  cofinite, 107
  discrete, 107
  finer, 108
  generated by a family of functions, 114
  generated by a family of sets, 108
  norm, 143
  order, 112
  product, 114
  quotient, 118
  relative, 107
  stronger, 108
  strong operator, 161
  trivial, 107
  of uniform convergence, 126
  vague, 217
  weak, 114, 161
  weaker, 108
  weak operator, 161
  weak*, 161
  Zariski, 111
Torus, 229
Totally bounded set, 15
Total ordering, 5
Total variation, of a complex measure, 88
  of a function, 96, 97
  of a signed measure, 82
Transfinite induction, 9
Transform, Fourier, 240, 260
  Fourier-Stieltjes, 283
  Hilbert, 266
Triangle inequality, 143
Trivial topology, 107
Tychonoff space, 117
Tychonoff's theorem, 130

Uncountable ordinal, first, 10
Uniform boundedness principle, 156
Uniform continuity, 14, 229, 313
Uniform convergence, topology of, 126
Uniformly integrable family, 87
Uniform norm, 115
Unimodular group, 320
Unitary isomorphism, 169
Upper bound, 5
Upper semi-continuous function, 211
Urysohn metrization theorem, 139
Urysohn's lemma, 116, 125, 237
U.s.c. function, 211

Vague topology, 217
Vanishing, of distributions, 263
  at infinity, 125
Variable, random, 286
Variance, 287, 293
  sample, 298
Variation, bounded, 97
Variation of a function, negative, 98
  positive, 98
  total, 96, 97
Variation of a measure, negative, 82
  positive, 82
  total, 82, 88
Vector space, normed, 143
  topological, 158
Vitali convergence theorem, 180
Vitali covering theorem, 105
Volume density, 333

Wave equation, 277
Weak convergence, 161
Weaker topology, 108
Weak law of large numbers, 294
Weak L^P space, 191
Weak operator topology, 161

Weak* topology, 161
Weak topology, 114, 161
Weak type, 195
Weierstrass approximation theorem, 135, 291
Weierstrass kernel, 246
Well ordered set, 5
Well ordering principle, 5
Weyl's lemma, 279
Wiener process, 306
  abstract, 305
Wirtinger's inequality, 247

Young's inequality, 232

Zariski topology, 111
Zorn's lemma, 5